# NETWORK MODELS AND INFORMATION DIFFUSION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Emily Marie Fischer

August 2019

NETWORK MODELS AND INFORMATION DIFFUSION

Emily Marie Fischer, Ph.D.

Cornell University 2019

The study of networks in such diverse areas as biology, technology, and social sciences has given rise to the interdisciplinary field of network science. Many real-world networks exhibit strongly connected communities and a degree distribution that follows a power law. This thesis explores these two topics - community structure and power law distributions - as they relate to network models and the diffusion of information on networks.

We first consider the generation of heavy-tailed distributions in stochastic processes. We give a system of stochastic differential equations in which processes grow at an exponential rate, but are reset at exponentially distributed times. We show that this system has a stationary solution which is regularly varying.

It is known that networks with a power law degree distribution are produced under the preferential attachment model, where edges are attached with preference to nodes of high degree. We analyze the effect of community structure on the degree distribution of a community-aware preferential attachment model. We also consider a generative network model where the metric for edge formation is not degree but the number of common neighbors.

We further study the effect of community structure on information diffusion in networks. Under the Susceptible-Infected-Susceptible model, we show that the epidemic threshold of a network is closely related to the epidemic threshold of its strongest community. We consider the lifetime of an infection on a grow-

ing preferential attachment network and show that the lifetime distribution has heavier tails on the growing network than on static networks.

## BIOGRAPHICAL SKETCH

Emily M. Fischer was born in Portland, Oregon. She received her Bachelor of Science in Mathematics from Harvey Mudd College in 2014 and began her PhD soon afterward at Cornell ORIE.

## ACKNOWLEDGEMENTS

I would like to thank my advisor, Gennady Samorodnitsky, as well as Souvik Ghosh and my other committee members. I thank Venus Lo for putting up with having me as an officemate for five years. Most importantly, I thank my husband, Diogo Magalhaes, for his support and encouragement throughout this entire process, and for making me coffee every morning.

## TABLE OF CONTENTS


Biographical Sketch . . . iii
Acknowledgements . . . iv
Table of Contents . . . v
List of Tables . . . vii
List of Figures . . . viii

**1 Introduction** **1**
1.1 Network Modeling . . . 2
1.1.1 Community Detection and the Stochastic Block Model . . . 5
1.1.2 Preferential Attachment Model . . . 8
1.2 Regular Variation . . . 12
1.3 Thesis Overview . . . 14

**2 A Two-Dimensional Poisson Counter Driven Stochastic Differential Equation** **17**
2.1 Introduction . . . 17
2.1.1 Preliminaries . . . 18
2.2 Model . . . 20
2.3 The Sampled Process at Reset Times . . . 22
2.4 The Joint Process . . . 31
2.5 Conclusion . . . 44

**3 Community-Aware Preferential Attachment Model** **45**
3.1 Introduction . . . 45
3.2 Model . . . 49
3.3 CAPAM with 2 Symmetric Communities . . . 53
3.3.1 Coupling $N_t(i, j)$ and $\hat{N}_t(i, j)$ . . . 58
3.3.2 Concentration of $\hat{N}_t(i, j)$ around the mean . . . 60
3.3.3 Expected degree distribution . . . 64
3.4 Regular Variation in the $k = 2$, symmetric case . . . 67
3.5 General $k$ . . . 71
3.5.1 Regular Variation for General $k$ . . . 72
3.5.2 Intuition for convergence, general $k$ . . . 75
3.6 Conclusion . . . 77

**4 Common Neighbors Model** **80**
4.1 Model . . . 81
4.2 K process . . . 84
4.2.1 Total Common Neighbors . . . 85
4.2.2 Common Neighbors between $i$ and $j$ . . . 93
4.3 Common Neighbors and Jaccard Similarity . . . 95
4.4 Modularity and Cluster Coefficient . . . 99
4.5 Conclusion . . . 103

**5 Information Diffusion on Networks** **104**
5.1 Introduction . . . 104
5.1.1 SIS Model . . . 106
5.1.2 Related Work . . . 109
5.2 SIS Model on Networks with Communities . . . 111
5.2.1 Main Result . . . 114
5.2.2 Numerical Example . . . 118
5.2.3 Conclusion . . . 121
5.3 SIS Model on Growing Networks . . . 123
5.3.1 Time-varying networks . . . 124
5.3.2 Model . . . 126
5.3.3 Evidence for infinite survival . . . 129
5.3.3.1 Mass at $\infty$ . . . 129
5.3.3.2 Conditional survival probability . . . 131
5.3.4 Lifetime distributions . . . 135
5.3.5 Epidemic Threshold . . . 137
5.3.6 Empirical epidemic threshold . . . 138
5.3.7 Infinite survival on the growing star graph . . . 140
5.3.8 Conclusion . . . 143

**A Coupling $G$ and $\hat{G}$** **145**
A.1 Coupling Overview . . . 145
A.2 Coupling Scheme . . . 147
A.2.1 Setup and Notation . . . 147
A.2.2 Detailed Procedure . . . 149
A.2.3 Subroutines . . . 151
A.3 Well-definedness . . . 157
A.3.1 By case in step 5(b) . . . 157
A.3.2 Marginal distributions . . . 159
A.4 Concentration Inequalities . . . 164
A.5 Proof of Lemma 5 . . . 177

**B Proof of Expected Degree Sequence** **184**

**Bibliography** **193**

## LIST OF TABLES

4.1 Collaboration graph properties . . . . . 102
4.2 Citation graph properties . . . . . 102

5.1 Summary of SIS notations . . . . . 139
5.2 Empirical epidemic threshold . . . . . 140

## LIST OF FIGURES


1.1 Protein-protein interaction network . . . . . 2
1.2 Sociology citation network . . . . . 4
1.3 Email network . . . . . 5
1.4 Example of SBM . . . . . 7
1.5 Example of a PA tree . . . . . 10
1.6 Example of a general PA graph . . . . . 11

2.1 Sample solution to the symmetric two-dimensional PCSDE . . . 22

3.1 CAPAM simulation with 2 symmetric communities . . . . . 71
3.2 CAPAM simulation with 3 communities . . . . . 75
3.3 Sample CAPAM graph . . . . . 78
3.4 Sample CAPAM graph with strong communities . . . . . 78

4.1 Average common neighbors and degree in the CNM . . . . . 96
4.2 Average common neighbors and degree in the CNM (labeled by node type) . . . . . 97
4.3 Low degree - high connectivity node . . . . . 97
4.4 Special cases of network connectivity . . . . . 98

5.1 Lifetime of SIS model on $G_1$ in the quick vs. slow extinction regimes . . . . . 108
5.2 Lifetime of SIS model on $G_2$ in the quick extinction regime . . . . 109
5.3 Comparisons of spectral radius and epidemic threshold for $G_1$, $G_2$, and $G$. . . . . 119
5.4 Histograms of extinction times on $G_1, G_2$, and $G$ with $\lambda = 0.18$ . . 120
5.5 Lifetime CCDFs for SIS on growing and static graphs . . . . . 129
5.6 Log-lifetime histograms of SIS on growing and static graphs . . . 130
5.7 Infection prevalence on a growing graph . . . . . 131
5.8 Empirical $P(\tau > n + 1000 \,|\, \tau > n)$ for SIS processes on (a) growing graph sequences and (b) a static graphs. . . . . 132
5.9 Empirical $P(\tau > n+100|\tau > n)$ for SIS processes with (top) $\lambda = 0.17$ on a growing graph and (bottom) $\lambda = 0.19$ on a static graph. Each curve corresponds to a distinct graph or graph sequence. . . . . 134
5.10 Lifetime tail distributions . . . . . 136

A.1 Tree diagram for coupling under $\tau_k = t > \hat{\tau}_k$ . . . . . 156

CHAPTER 1

# INTRODUCTION

The study of networks in such diverse areas as biology, technology, and social sciences has given rise to the interdisciplinary field of network science. Recent efforts have been made to discern the organization and relationships of complex networks, understand and predict network formation, and optimize network processes. Many real-world networks exhibit a complex structure and irregular connectivity patterns. At the same time, several network properties arise in profusion. These include a configuration of strongly connected communities and the manifestation of heavy tails in the degree distribution. This thesis explores these two topics - community structure and power law distributions - as they relate to network models and the diffusion of information on networks.

In the coming pages, we consider the mechanisms by which heavy tails are generated in network models, epidemic spreading, and other stochastic processes. We theorize on the formation of communities and discuss the implications of community structure on network growth and information diffusion. But before we delve into the new work, it is important to give some background on networks. In Section 1.1, we discuss typical properties of real-world networks and present two well-known network models. In Section 1.2, we discuss the concept of regular variation. Section 1.3 ends this chapter with an overview of the thesis.

## 1.1 Network Modeling

Networks, generally defined as systems of interacting entities, are found in a wide variety of application areas, from computer networks and social networks, to ecological networks and neural networks. We often represent networks as mathematical graph, where nodes represent actors or individuals of some population, and edges represent interaction between two individuals.

A social network is a network of interpersonal interactions among a population, where such interactions may occur in person or online, in the form of an exchange of information or media. The study of social networks is vital to understanding dissemination of information and the transmission of infectious disease. For instance, social network analysis has been employed to study the spread and prevention of HIV [37, 49, 76, 54, 10]. In the last couple of decades, the advent of online social networks like Facebook, LinkedIn, and Twitter have created a wealth of data, allowing researchers to study the structure and activity of large, complex networks like never before [45, 67, 29, 18].

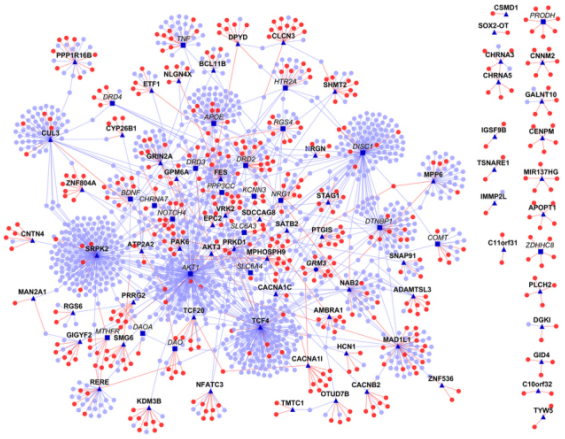



Figure 1.1: Protein-protein interaction network for genes associated with schizophrenia [32].

It is often useful to model large, complex networks using random graphs. The simplest random graph model is the Erdős-Rényi (ER) random graph, which consists of a fixed number of nodes, and each possible edge is present with probability $p$, independently from all other edges [39].

The ER random graph is appealing in its simplicity, but it lacks certain properties that are found in many real-world networks. For instance, the distribution of the degree of any node in an ER random graph is binomial, and as the number of nodes $n \to \infty$, the degree distribution converges to a normal distribution. Thus the asymptotic graph has exponentially decaying tails. Meanwhile, many real networks exhibit the so-called *scale-free property*, meaning that their degree distributions decay as a power law: the fraction of nodes with degree $k$ decays like $k^{-\gamma}$ for some positive power $\gamma$. This is a heavy-tailed distribution, and appears in practice when a relatively small fraction of the nodes spawn a relatively large fraction of the edges.

Having observed this phenomenon, Barabási and Albert proposed the *preferential attachment* model (PAM) to generate graphs with a power law degree distribution [8]. The PAM is based off of the idea that new edges are more likely to be attached to higher-degree nodes. This model has become widely used over the last several years, and given rise to many variants. We discuss this model in more detail in Section 1.1.2.

Another common property of real-world networks is the presence of *communities*: groups of densely interconnected nodes, with relatively few connections between groups [31]. In many applications, communities correspond to different types or categories of nodes. For instance, in the sociology citation network in Figure 1.2, the detected communities appear to correspond to certain topics

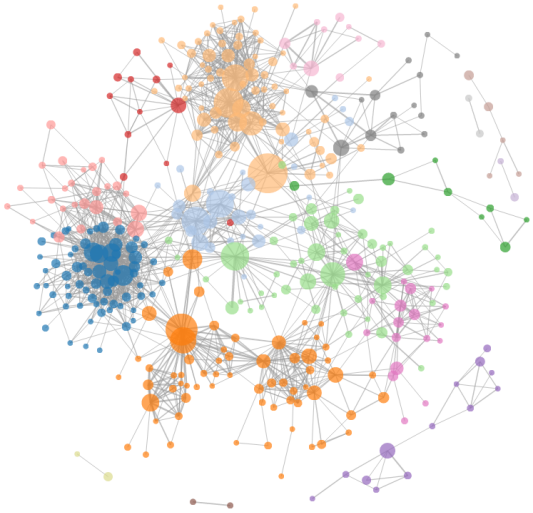


Figure 1.2: A sociology citation network due to [17]. Node colors differentiate communities, as detected by a community detection algorithm.

or subfields of sociology [17].

The Stochastic Block Model (SBM) is commonly used to model graphs with communities. It is similar to the ER model, but each node has an assigned community membership, and the probability of forming an edge between two nodes depends on their community membership [2]. Community structure is produced under the assumption that edges are more likely to form between two nodes of the same community than they are between nodes of different communities. The SBM is discussed further in Section 1.1.1, along with its relation to community detection.

There are numerous variants of PAM and SBM, as well as other types of random graph models. We introduce two new random graph models in this thesis. The first is the Common Neighbors Model (CNM), in which edges tend to form between nodes with a large number of common neighbors. This model is described in detail in Chapter 4. The second is a Community-Aware Preferential Attachment Model, and is defined and rigorously analyzed in Chapter 3.

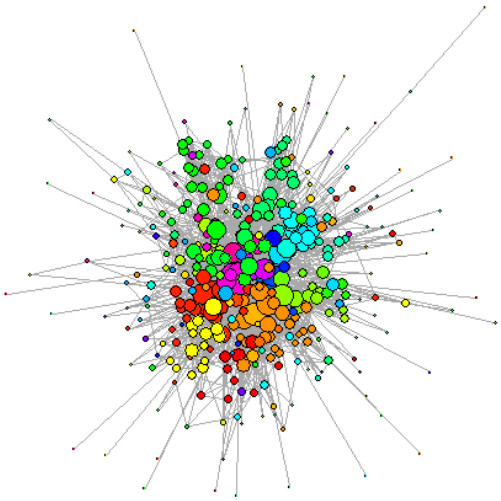


Figure 1.3: Email network of a research institute with 42 departments. A total of 361 nodes and 3699 edges are pictured, where node color corresponds to departmental membership and node size is proportional to the square root of its degree.

### 1.1.1 Community Detection and the Stochastic Block Model

There are many examples of real-world networks with communities. In some networks, communities are inherent and correspond to certain traits or features. For example, Figure 1.3 shows an email network with communities corresponding to departments of a European research institute [92, 56].

Not all networks come labeled according to community membership, however. A huge field of research is devoted to *community detection*, the process of identifying the communities in an unlabeled graph. Many comunity detection algorithms operate by partitioning a graph in a way that optimizes some measure of community structure. Unfortunately, though, there is no universally agreed-upon definition of community. Though the notion of community may be intuitively clear, there is no obvious mathematical definition. This leads to vari-

ous interpretations of community, and numerous measures of the "strength" of community structure in a graph. Each measure, or score, of community strength then defines its own class of community detection algorithms [31, 72].

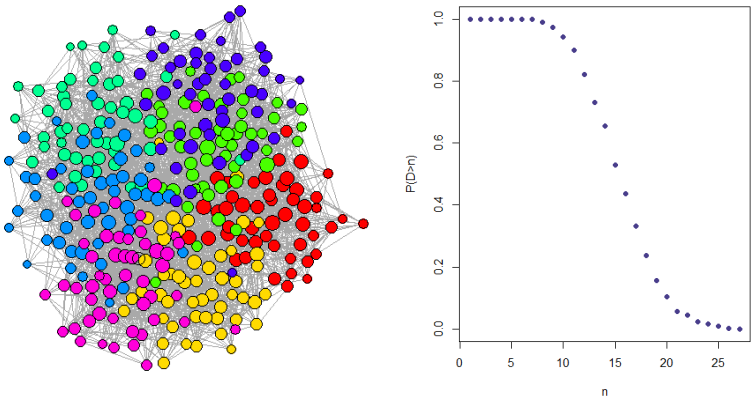
The concept of *modularity*, introduced in Girvan and Newman, is perhaps the most well-known measure of community structure, and is defined as

$$Q = \frac{1}{2m} \sum_{i,j} \left( A_{ij} - \frac{d_i d_j}{2m} \right) \delta(C_i, C_j), \tag{1.1}$$

where $m$ is the total number of nodes in the graph, $A$ is the adjacency matrix, $d_i$ and $d_j$ are the degrees of nodes $i$ and $j$, respectively, and $\delta(C_i, C_j)$ is equal to 1 if $i$ and $j$ belong to the same community, and 0 otherwise [35]. This quantity calculates the proportion of edges connecting nodes of the same community, and subtracts the expected proportion in a graph with the same node degrees, but with edges connected at random. Thus, the higher the modularity, the stronger the communities.

Unfortunately, however, modularity optimization does not always result in a desirable community partition. Fortunato and Barthélemy point out that a resolution limit exists when modularity optimization is used for community detection. The maximum value of modularity changes with graph size, and the maximum may be attained by a graph partition that merges unmistakably distinct communities [30]. Thus the output of a modularity-optimizing algorithm may not be credible. Community detection algorithms must be thoroughly evaluated for accuracy and reliability, urge Yang and Leskovic, who evaluate the performance of several community measures by studying 230 large real-world networks [91].

Community detection algorithms can be theoretically evaluated on the Stochastic Block Model. As the SBM generates graphs with ground-truth com-




(a) (b)

Figure 1.4: (a) Example of SBM output on 350 nodes and 7 communities, with edge probabilities $Q_{ii} = 0.2$, $i = 1, \ldots, 7$, and $Q_{ij} = 0.02$ for all $i \neq j$. Node size is proportional to the square root of degree. (b) Complementary CDF of node degree in this SBM graph.

munities, detection algorithms can be evaluated based on the proportion of nodes correctly recovered. Since the SBM has a relatively simple, Erdős-Rényi-like structure, it allows for theoretical guarantees on recovery [25, 47, 93, 2].

Formally, the SBM is defined by the following parameters: a number of nodes $n$; a number of communities $k$, a probability vector $p = (p_1, \ldots, p_k)$, and a $k \times k$ matrix $Q$ with entries in $[0, 1]$. Each of the $n$ nodes is randomly assigned a community membership in $\{1, \ldots, k\}$, according to the probability vector $p$. An edge is formed between two nodes with probability $Q_{ij}$, where $i, j \in \{1, \ldots, k\}$ are their respective community memberships.

The SBM has been around almost as long as the ER model [40], and has been also been called the planted partition model [15, 28] and an inhomogeneous random graph [13]. Recent work in community detection focuses on determining the parameter regimes in which full or partial recover is possible. There are other problems of interest, however, including estimating model parame-

ters based on a single graph realization.

The advantage of the SBM is that one can specify the parameters to customize the community structure. However, like the ER model, the degree distribution of a node in the SBM is bounded by an exponential distribution. Therefore the SBM may not be an ideal model for many real networks. For example, consider the realization of the SBM depicted in Figure 1.4 in comparison to the email network Figure 1.3. The SBM has a more uniform structure, and less variance in degree.

Some authors have proposed a degree-corrected SBM that induces a power law degree distribution [47, 74]. Others work in the other direction, introducing community structure into random graph models, such as the Preferential Attachment Model, that produce power law degree distributions. For instance, Hajek and Sankagiri modifiy the PAM to include a fitness function based on community membership [38]. Our CAPAM model is a novel variant of the PAM that incorporates community structure via edge acceptance probabilities (see Chapter 3). The Preferential Attachment Model is the subject of the next section.

### 1.1.2 Preferential Attachment Model

Many real networks are known to follow power law distribution, including citation networks, url links on the world wide web, and others [6, 27, 51, 66]. As an attempt to explain the source of this phenomenon, Barabási and Albert proposed a random graph generated by a *preferential attachment mechanism* that would result in a power law degree distribution [8]. This was formalized soon

after by by Bollobás et al. [14].

The preferential attachment model is a generative random graph model, meaning that the graphs are grown over time by a set of recursive rules. The growth rules are based off of the fundamental assumptions that nodes join a network at a constant rate and that they tend to form connections with nodes that already have many connections. Thus, the higher a node's degree, the more likely it is to increase its degree. Over time, this rich-get-richer effect results in a power law degree distribution [8].

More specifically, the model initates a directed graph $G_0$ with a single node and zero edges. Given the graph $G_t$ at some time $t \geq 0$, $G_{t+1}$ is formed by adding the following to $G_t$:

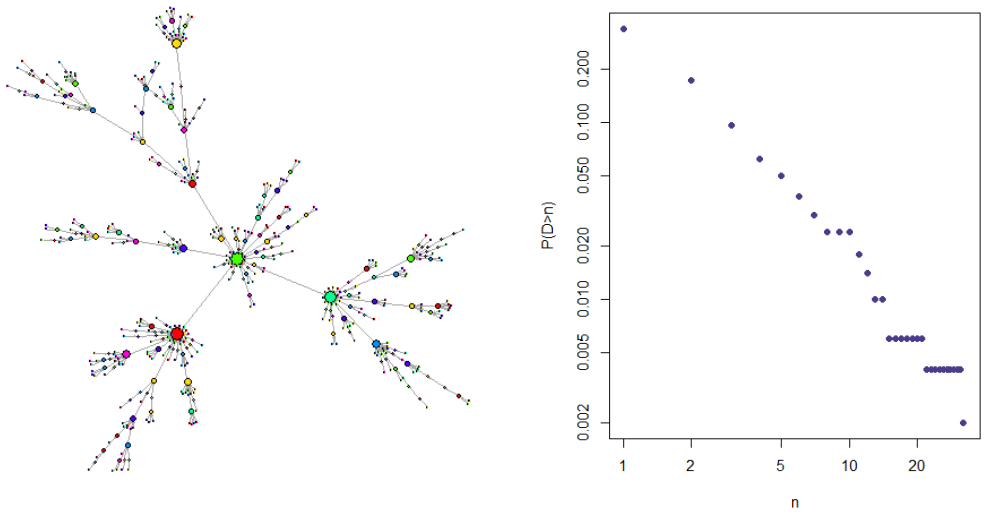
- a new node $v$
- the directed edge $(v, w)$, where $w$ is selected with probability

$$\frac{D_t(w) + \delta}{t + \delta(t+1)},$$

where $\delta > 0$ is a constant parameter and $D_t(w)$ denotes the in-degree of node $w$ at time $t$.

Under this formulation, the resulting graph is a tree (see Figure 1.5 for an example). This basic model is made more general by allowing for $m \geq 1$ edges to be formed at each time step [39] or by allowing new edges to form between existing nodes [14]. See, for example, Figure 1.6.

As $t \to \infty$, the generated graph approaches the so-called *asymptotic graph*, which is shown to follow a power law in-degree distribution. Specifically, for $G_t$ generated according to the basic rules above, we have the following theorem.




(a) (b)

Figure 1.5: (a) Example of a PA tree on 500 nodes. Nodes are pictured with size proportional to square root of their degree. (b) Complementary CDF of node degree in this PA tree, in log-log scale.

**Theorem 1.** *Let $P_k(t)$ denote the proportion of nodes with in-degree $k$ in $G_t$. Then for any fixed $k \geq 0$,*

$$P_k(t) \to p_k$$

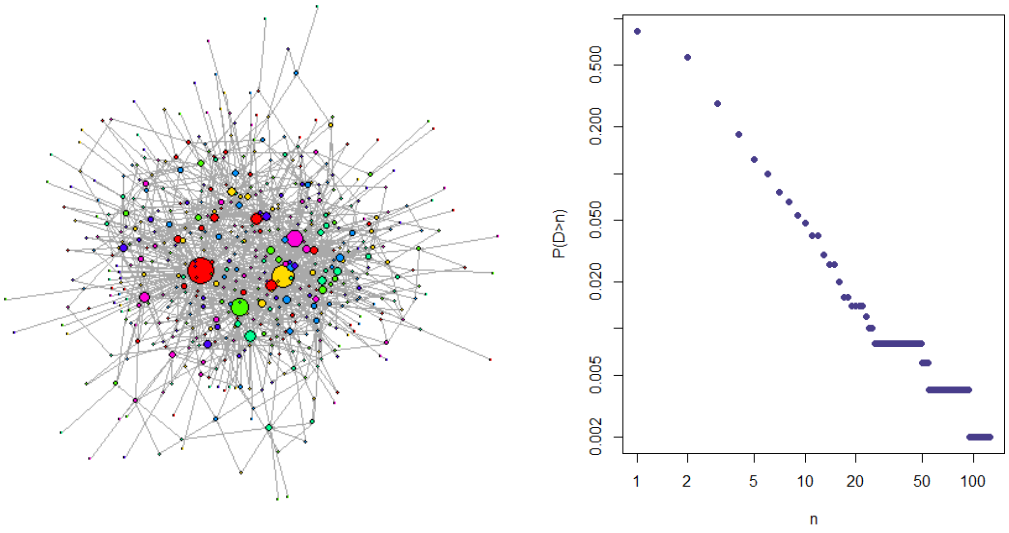
*as $t \to \infty$, where*

$$p_k = (2+\delta)\frac{\Gamma(k+\delta)\Gamma(1+2\delta)}{\Gamma(k+2+2\delta)\Gamma(\delta)}, \qquad k \geq 0.$$

Due to the ratio of gamma functions, we see that $p_k$ converges to a power law for large $k$, with $p_k \sim k^{-(2+\delta)}$. In Chapter 3, we will see that the asymptotic distribution $(p_k, k \geq 0)$ is regularly varying, as in Definition 2 on page 12.

For the proof of this theorem and a description of a more general PAM, we refer the reader to [39].

There is an abundance of variations on the PAM. Cooper and Frieze propose a general preferential attachment model where out-degree $M_t$ is chosen according to some probability distribution independently at each time step [23]. The




(a) (b)

Figure 1.6: (a) Example of a PA graph on 500 nodes with out-degree $m$ taking values in $\{1, 2, 3\}$. Nodes are pictured with size proportional to square root of their degree. (b) Complementary CDF of node degree in this PA graph, in log-log scale

model can be altered so that out- and in-degrees are treated differently, and the joint degree distribution is analyzed in [79].

The standard PAM is sometimes called linear preferential attachment, whereas a non-linear preferential attachment model depends in a nonlinear way on node degree. Krapivsky et al. consider the case where the probability of attaching to a node with degree $d$ is proportional to $d^\alpha$ for some power $\alpha$ [50]. The superlinear ($\alpha > 1$) case is studied in detail in [65].

We found the literature to be lacking, however, in a generative random graph model that generated community structure as well as a power law degree distribution. This was the intuition for our Community-Aware Preferential Attachment Model (CAPAM), which incorporates communities into the PAM by adding a response step. The response step models the connection process of

online social networks like Facebook and LinkedIn, where one user sends an invitation to another, and then the second user responds by chosing whether to accept the invitation or not. In the CAPAM, we assume that the probability of accepting an edge invitation is dependent upon the community membership of the two nodes in question. We will see in Chapter 3 how this model allows the user to specify community structure, and that the resulting graph remains scale free.

We must also note that a recent model due to Hajek and Sankagiri, which incorporates communities in the PAM by using a multiplicative fitness that is dependent on community membership [38]. This is a special case of the geometric preferential attachment model given in [46]. Though this model is similar to CAPAM, we give a stronger result, specifying the multivariate degree distribution, instead of merely the marginal distributions. Further, we prove multivariate regular variation, as in Definition 2, and specify the spectral measure. Our methodology uses a novel coupling proof, and our use of edge acceptance probabilities has unique implications for networks that operate according to the edge invitation and response steps.

## 1.2 Regular Variation

In this section we define and discuss *regular variation*. Regular variation is the formal concept we will use to describe probability distributions with tails that look like a power law.

**Definition 2.** A nonnegative random variable $X$, or its distribution $F$, is called *regularly varying* if $\overline{F}(x) > 0$ for all $x \geq 0$, and there is an $\alpha \geq 0$ such that for every

$b > 0$,

$$\lim_{x\to\infty} \frac{\overline{F}(bx)}{\overline{F}(x)} = b^{-\alpha}. \tag{1.2}$$

Such an $\alpha$ is called the *tail exponent*.

Importantly, the tails of such a regularly varying $X$ follow a power-law distribution with exponent $\alpha$.

We note that any function that decays as a pure power law, e.g., some $f$ where $\lim_{x\to\infty} x^{\alpha} f(x) = 1$, is necessarily regularly varying. However, regular variation is a slight generalization to the former statement, as can be seen by the example $g(x) = x^{-\alpha}\log(x)$. In this example, $\lim_{x\to\infty} x^{\alpha} g(x) = \infty$, however

$$\lim_{x\to\infty} \frac{g(bx)}{g(x)} = \lim_{x\to\infty} \frac{(bx)^{-\alpha}\log(bx)}{x^{-\alpha}\log(x)} = b^{-\alpha}.$$

Thus $g$ meets the conditions of Definition 2, and is regularly varying.

More generally, any regularly varying $X$ can be written as the product of a power law and a *slowly varying* function. A function $L$ is called slowly varying if it is regularly varying with tail exponent equal to 0. Therefore a random variable $X$ is regularly varying with tail exponent $\alpha$ if and only if

$$P(X > x) = x^{-\alpha} L(x), \quad x \geq 0 \tag{1.3}$$

for some slowly varying function $L$. .

We can further extend regular variation to the multivariate case.

**Definition 3.** A random vector $\boldsymbol{X} \in \mathbb{R}^d$, or its law, is said to be regularly varying if $P(\|\boldsymbol{X}\| > x) > 0$ for all $x \geq 0$, and there exist $\alpha \geq 0$ and a probability measure $\sigma$ on the unit sphere $\mathbb{S}^{d-1}$ of $\mathbb{R}^d$ such that for every $b > 0$,

$$\lim_{x\to\infty} \frac{P(\|\boldsymbol{X}\| > bx,\ \boldsymbol{X}/\|\boldsymbol{X}\| \in \cdot)}{P(\|\boldsymbol{X}\| > x)} = b^{-\alpha}\sigma(\cdot) \tag{1.4}$$

vaguely in $\mathbb{S}^{d-1}$.

Definition 3 implies that a regularly varying multivariate distribution is characterized by two quantities: the tail index $\alpha$ and the *spectral measure* $\sigma$. The tail exponent indicates the decay rate of the size of the vector, while the spectral measure gives a distribution on the direction of the vector.

An alternate definition in the multivariate case is as follows. We say that a random vector $\mathbf{X}$ is regularly varying if there exists a Radon measure $\nu$ and a function $b(t) \to \infty$ such that as $t \to \infty$,

$$tP\left(\frac{\mathbf{X}}{b(t)} \in \cdot\right) \xrightarrow{v} \nu(\cdot), \tag{1.5}$$

where $\xrightarrow{v}$ denotes vague convergence.

We will use multivariate regular variation in Chapters 2 and 3 to describe the joint distributions of two processes of interest.

## 1.3 Thesis Overview

In Chapter 2, we take a short departure from networks to consider system of stochastic differential equations which yield heavy-tailed solutions. This work illustrates that power laws result when exponentially growing processes are reset at exponentially distributed times. We examine a two-dimensional stochastic differential equation where each process grows exponentially, but is reset to 1 at arrivals of an independent Poisson process. We show that this process has a stationary solution which is regularly varying.

In Chapter 3, we return to the question of network modeling, and propose

a Community-Aware Preferential Attachment Model (CAPAM). This model extends the traditional PAM by replacing the edge formation step with a two-step process: invitation and response. In the CAPAM, the initially selected node $v$ sends an edge invitation to another node, $w$. Then $w$ responds, either by accepting the invitation and forming the edge $(v, w)$, or by rejecting the invitation, in which case no edge is formed. We assume that the probability of a positive response is dependent on the community memberships of the two associated nodes. One can specify the strength of communities in this model through the matrix of acceptance probabilities.

An in-depth coupling proof is used to determine the asymptotic degree distribution of this model, where degree is a vector denoting the number of edges pointing to each community. We show that the degree distribution is multivariate regularly varying, with spectral measure that depends on the acceptance probability matrix. Finally, we present a few simulations and discuss the implications of the invitation-response process in network models.

In the CAPAM, nodes have an inborn community membership, and community structure is formed in the graph through community-dependent edge acceptance probabilities. Not all real-world networks have ground-truth communities, however, and it may be desirable to create community structure without having to designate a matrix of community parameters, such as is the case in the CAPAM and the SBM.

In Chapter 4, we propose the Common Neighbors Model (CNM) to do just this. In the CNM, edges are attached preferentially to nodes sharing a high number of common neighbors. Thus, these edges tend to close triangles and create dense clusters of connected nodes. We give simulation results illustrating the

structure of graphs produce through this model. Although we have found analysis of the degree distribution to be intractable, we introduce a method to separate the evolution of the graph from the evolution of the number of common neighbors of each pair of nodes. We give a model for the so-called $K$ process for common neighbors, and give partial analytical results. We further examine common neighbors in LinkedIn data as well as our CNM simulations.

Chapter 5 shifts to the topic of information diffusion on networks. We discuss how the spread of information on networks can be studied with established epidemic models. In Section 5.1, we give a summary of the Susceptible-Infected-Susceptible (SIS) epidemic model and important results. We then consider the SIS model in two distinct contexts. In Section 5.2, we look at the behavior of the SIS model on networks with communities. We give a result on the epidemic threshold of networks with communities and illustrate this through a numerical example. In Section 5.3, we study the propagation of information on prefential attachment networks that are growing over time. We perform an in-depth simulation study and observe that the lifetime distribution of the SIS model has a heavier tail on growing networks than on static networks. We show that, in fact, infinite survival may be possible on a growing preferential attachment network and we give a short proof of the possibility of infinite survival on a growing star network. We conclude by identifying open questions and directions for future work.

CHAPTER 2

# A TWO-DIMENSIONAL POISSON COUNTER DRIVEN STOCHASTIC DIFFERENTIAL EQUATION

Heavy-tailed distributions have been observed in a variety of phenomena, and the origin of such behavior has been attributed to the reset of an exponentially growing process at exponential times. We present a class of 2-dimensional Poisson Counter Driven Stochastic Differential Equations (PCSDE) that give rise to a bivariate power law distribution. We prove the existence of a stationary distribution to such equations through basic principles of stochastic stability. Our main result proves multivariate regular variation and examines the tail dependence.

## 2.1 Introduction

A variety of naturally occuring phenomena have been observed to follow a heavy-tailed distribution, including household wealth, the number of url links to web pages and even the size of forest fires [20]. Therefore there is interest in how such distributions arise through natural processes. It is shown in [75] that a heavy-tailed distribution emerges when an exponentially growing process is observed and reset at exponentially distributed random times.

Along this line, the authors in [59] propose several 2-dimensional *Poisson Counter Driven Stochastic Differential Equation* (PCSDE) models. These models are stochastic differential equations that describe a pair of dependent processes that grow exponentially, but are stopped and renewed at arrivals of independent Poisson processes. This has significance to, for example, the joint in- and

out- degree distribution of the Preferential Attachment model [79].

The aim of the current work is to be a rigorous extension of the previous paper, explicitly proving the existence of a heavy-tailed stationary distribution to a proposed stochastic differential equation. We examine a two-dimensional PCSDE model $(X_1(t), X_2(t))_{t\geq 0}$ in which the two processes are coupled symmetrically. Each process grows exponentially according to a deterministic differential equation, but each process is renewed to a value of 1 at arrivals independent Poisson processes.

This model yields a stationary distribution that is regularly varying. We additionally consider the tail dependence coefficient

$$\chi = \lim_{t\to\infty} P(X_2 > t | X_1 > t),$$

a measure that describes how the two processes, $X_1$ and $X_2$, depend on each other at extreme values [22]. The tail dependence coefficient of our processes is strictly positive, and hence the two processes are asymptotically dependent.

On first attempt, we endeavored to find a drift condition that would prove stationarity of the solution to this model. However, such a drift condition was only successful over a strict subset of the parameter space. Hence we prove stationarity using baser principles. Once stationarity is proved, regular variation follows from the structure of the model.

### 2.1.1 Preliminaries

In this subsection we present a few concepts that will be essential throughout the chapter. The first important concept is regular variation, as defined in Sec-

tion 1.2. A regularly varying distribution has tails that decay as a power law according to the tail exponent $\alpha$. We refer to multivariate regular variation, as in Definition 3, when discussing the solutions to the Poisson counter driven stochastic differential equation.

In Section 2.3 we refer to the concept of a *petite set* for Markov chains. This idea is based on *samplings* of the original process, in which the process is observed at random times following a distribution $a$. For the rest of this section, suppose we have a Markov process $X = (X_t; t \geq 0)$ in $\mathbb{R}^d$ with transition kernel $P^t(x, \cdot)$. All of the following definitions are from [61].

**Definition 4.** Let $a$ be a probability measure on $\mathbb{R}_+$. Define Markov transition function $K_a$ as

$$K_a := \int P^t a(dt).$$

The Markov chain associated with transition function $K_a$ is called the $K_a$-chain and is a sampled chain of the original process.

In many cases, a sampled chain retains the same recurrence properties as the orginal process. This fact will be used on the chain resulting from sampling our process at renewal times .

The $K_a$-chain is inherent to the idea of petiteness of sets. A set is petite if, for $x$ in the set, the probability measure $K_a(x, \cdot)$ dominates some other measure of the state space.

**Definition 5.** Let $a$ be a probability measure on $\mathbb{R}_+$. A Borel set $C \in \mathcal{B}(\mathbb{R}^d)$ is called $\nu_a$-*petite* if there exists a non-trivial measure $\nu_a$ on $\mathcal{B}(\mathbb{R}^d)$ such that $K_a(x, \cdot) \geq \nu_a(\cdot)$ for all $x \in C$.

We use petite sets to prove Harris recurrence of both the original process and a sampled chain. To prove *positive* Harris recurrence of the process, we additionally show that the process is *bounded in probability on average*.

**Definition 6.** The process $X$ is called bounded in probability on average if for any initial condition $x \in \mathbb{R}^d$ and each $\epsilon > 0$, there exists a compact subset $C \subset \mathbb{R}^d$ such that

$$\liminf_{t\to\infty} \frac{1}{t} \int_0^t P_x(X_s \in C) ds \geq 1 - \epsilon.$$

The last definition is that of a *T-process*, a process in which positive Harris recurrence and boundedness in probability in average are equivalent.

**Definition 7.** A kernel $T$ is called a *continuous component* of a function $K : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \to \mathbb{R}_+$ if

(a) For $A \in \mathcal{B}(\mathbb{R}^d)$, the function $T(\cdot, A)$ is lower semi-continuous, and

(b) For all $x \in \mathbb{R}^d$ and $A \in \mathcal{B}(\mathbb{R}^d)$, the measure $T(x, \cdot)$ satisfies $K(x, A) \geq T(x, A)$.

A process is a *T-process* if for some $a$, the $K_a$-chain admits a continuous component $T$ which is non-trivial for all $x \in \mathbb{R}^d$, that is, $T(x, \mathbb{R}^d) > 0$, $x \in \mathbb{R}^d$.

## 2.2 Model

In the most general case, our model is governed by the following system of stochastic differential equations, with state space $[1, \infty) \times [1, \infty)$.

$$\begin{aligned} dX_1(t) &= \mu_{11} X_1(t) dt \ + \ \mu_{12} X_2(t) dt \ + \ (1 - X_1(t^-)) dN_1(t) \\ dX_2(t) &= \mu_{21} X_1(t) dt \ + \ \mu_{22} X_2(t) dt \ + \ (1 - X_2(t^-)) dN_2(t). \end{aligned} \tag{2.1}$$

Thus $X_1$ and $X_2$ both grow exponentially fast, depending on the parameters $\mu_{ij}$, $i, j = 1, 2$, but are reset to 1 at each arrival of the corresponding Poisson processes $N_1$ and $N_2$, with rates $\lambda_1$ and $\lambda_2$, respectively.

For simplicity, we consider the symmetric problem

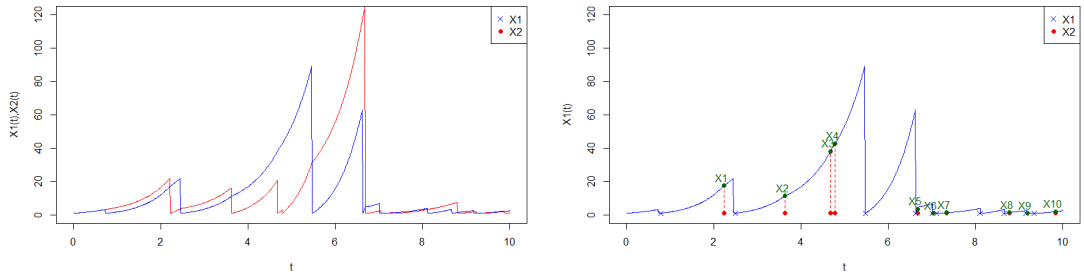
$$\begin{aligned} dX_1(t) &= X_1(t)dt + \mu X_2(t)dt + (1 - X_1(t^-))dN_1(t) \\ dX_2(t) &= \mu X_1(t)dt + X_2(t)dt + (1 - X_2(t^-))dN_2(t), \end{aligned} \tag{2.2}$$

with Poisson rates $\lambda_1 = \lambda_2 = \lambda$. Here, $X_1$ and $X_2$ are symmetric processes, and their exponential growth rates depend only on the parameter $\mu$. See Figure 2.1a for an illustration of a possible solution to this problem.

Note that between arrivals of the Poisson process, the growth rates are completely deterministic. In particular, under the assumption that there are no arrivals in the interval $[0, t]$, we can calculate the values of $X_1(t)$ and $X_2(t)$ according to their initial values, as follows.

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} e^{t(1+\mu)} + e^{t(1-\mu)} \\ e^{t(1+\mu)} - e^{t(1-\mu)} \end{pmatrix} X_1(0) + \frac{1}{2}\begin{pmatrix} e^{t(1+\mu)} - e^{t(1-\mu)} \\ e^{t(1+\mu)} + e^{t(1-\mu)} \end{pmatrix} X_2(0). \tag{2.3}$$

In section (2.3), we sample the solution to (2.2) at reset times of $X_2$ to form a discrete-time, one-dimensional process. We show that this sampled process has a stationary distribution, and that this stationary distribution is regularly varying according to Definition 2. In section (2.4), we formulate the continuous, two-dimensional process in terms of the sampled process, in order to prove the analogous results. Indeed, the two-dimensional process has a stationary distribution which satisfies (1.4) for bivariate regular variation. We conclude by calculating the tail dependence coefficient.




(a) (b)

Figure 2.1: (a) A sample solution to (2.2). For the same example, (b) depicts the arrivals of $N_2$ (red) and the value $X_n$ at each arrival (green).

## 2.3 The Sampled Process at Reset Times

In this section we examine the Markov Chain consisting of the values of $X_1(t)$ at arrivals of the *second* Poisson process, $N_2$. This sampled process will exemplify the recurrence and regular variation properties we hope to find in the original process, and in fact the existence of these properties in the sampled process necessitates their existence in the original process.

Define $X_n$ to be the value of $X_1(t)$ at the $n^{\text{th}}$ arrival of $N_2$. This is illustrated in Figure 2.1. We consider the Markov Chain $(X_n; n \geq 0)$. The first lemma shows that $(X_n)$ follows a stochastic recursion. From there, we prove positive recurrence and regular variation of $(X_n)$.

**Lemma 1.** *The process $(X_n; n \geq 0)$ follows a stochastic recurrence of the form*

$$X_{n+1} = A_{n+1}X_n + B_{n+1}, \quad n = 0, 1, \ldots \tag{2.4}$$

*where $(A_n, B_n), n = 1, 2, \ldots$ are i.i.d. and $(A_{n+1}, B_{n+1})$ is independent of $X_n$.*

*Proof.* We fix two consecutive arrivals of $N_2$, and then consider the arrivals of $N_1$ that occur within this interval. Fix $n \geq 0$ and let $E_n$ be the time of the $n^{\text{th}}$ arrival

of $N_2$. Let $K$ be the number of arrivals of $N_1$ in the interval $(E_n, E_{n+1})$. Then let $T_1, \ldots, T_K$ be the times of arrivals of $N_1$ such that $E_n < T_1 < T_2 < \cdots < T_K < E_{n+1}$.

The time point $E_n$ is a sampling time for $X_1$, and a reset time for $X_2$. So at time $E_n$, $X_1$ and $X_2$ take on the values

$$X_1(E_n) = X_n, \quad X_2(E_n) = 1.$$

To simplify notation, we denote by $Y_j$ the value of $X_2$ at the $j^{\text{th}}$ arrival of $N_1$. That is, $Y_j = X_2(T_j)$. Since there are no arrivals of either process between $E_n$ and $T_1$, we can use the deterministic expression for $X_2$ in (2.3) to find an expression for $Y_1$.

$$\begin{aligned} Y_1 &= X_1(E_n)\frac{e^{(T_1-E_n)(1+\mu)} - e^{(T_1-E_n)(1-\mu)}}{2} + X_2(E_n)\frac{e^{(T_1-E_n)(1+\mu)} + e^{(T_1-E_n)(1-\mu)}}{2} \\ &= X_n\frac{e^{(T_1-E_n)(1+\mu)} - e^{(T_1-E_n)(1-\mu)}}{2} + \frac{e^{(T_1-E_n)(1+\mu)} + e^{(T_1-E_n)(1-\mu)}}{2} \end{aligned}$$

Next we find $Y_j$ for $j = 2, \ldots, K$. At time $T_{j-1}$, we have $X_1(T_{j-1}) = 1$, $X_2(T_{j-1}) = Y_{j-1}$. Therefore (2.3) yields

$$Y_j = \frac{e^{(T_j-T_{j-1})(1+\mu)} - e^{(T_j-T_{j-1})(1-\mu)}}{2} + Y_{j-1}\frac{e^{(T_j-T_{j-1})(1+\mu)} + e^{(T_j-T_{j-1})(1-\mu)}}{2}, \quad j = 2, \ldots, K.$$

Solving this recursion, we write $Y_K$ in terms of only $X_n$.

$$\begin{aligned} Y_K = &\sum_{i=1}^{K}\left[\left(\frac{e^{(T_i-T_{i-1})(1+\mu)} - e^{(T_i-T_{i-1})(1-\mu)}}{2}\right)\prod_{j=i+1}^{K}\left(\frac{e^{(T_j-T_{j-1})(1+\mu)} + e^{(T_j-T_{j-1})(1-\mu)}}{2}\right)\right] \\ &+ X_n\left(\frac{e^{(T_1-E_n)(1+\mu)} - e^{(T_1-E_n)(1-\mu)}}{2}\right)\prod_{j=2}^{K}\left(\frac{e^{(T_j-T_{j-1})(1+\mu)} + e^{(T_j-T_{j-1})(1-\mu)}}{2}\right) \\ &+ \prod_{j=1}^{K}\left(\frac{e^{(T_j-T_{j-1})(1+\mu)} + e^{(T_j-T_{j-1})(1-\mu)}}{2}\right). \end{aligned} \tag{2.5}$$

Finally, to determine $X_{n+1}$, we must consider the time between $T_K$ and $E_{n+1}$. At time $T_K$, we have $(X_1, X_2) = (1, Y_K)$, so

$$X_{n+1} = \frac{e^{(E_{n+1}-T_K)(1+\mu)} + e^{(E_{n+1}-T_K)(1-\mu)}}{2} + Y_K \frac{e^{(E_{n+1}-T_K)(1+\mu)} - e^{(E_{n+1}-T_K)(1-\mu)}}{2}. \tag{2.6}$$

Putting (2.5) and (2.6) together, we see that $X_{n+1} = AX_n + B$ for two random variables $A$ and $B$. Letting $F_1 = T_1 - E_n$, $F_{K+1} = E_{n+1} - T_K$, and $F_j = T_j - T_{j-1}$ for $j = 2, \ldots, K$, we can write an expression for $A$ as follows. If $K = 0$, then

$$A = \frac{1}{2}\left(e^{F_1(1+\mu)} + e^{F_1(1-\mu)}\right),$$

and if $K \geq 1$,

$$A = \frac{e^{F_1(1+\mu)} - e^{F_1(1-\mu)}}{2} \cdot \frac{e^{F_{K+1}(1+\mu)} - e^{F_{K+1}(1-\mu)}}{2} \cdot \prod_{j=2}^{K} \frac{e^{F_j(1+\mu)} + e^{F_j(1-\mu)}}{2},$$

where we consider the empty product to be 1. The random variable $B$ can similarly be expressed in terms of $F_1, F_2, \ldots, F_{K+1}$, and $K$ (see proof of Lemma 2).

Observe that $F_1, \ldots F_{K+1}$ are independent exponential random variables, and independent of $X_n$, each with parameter $2\lambda$. Furthermore, $K$ is a geometric random variable with probability of success equal to the probability that the next arrival will come from $N_2$ (instead of $N_1$). So $K$ is also independent of $X_n$, and therefore $A$ and $B$ are independent from $X_n$.

Hence we can write for each $n = 0, 1, \ldots$ that $X_{n+1} = A_{n+1}X_n + B_{n+1}$, where $A_{n+1}$ and $B_{n+1}$ are independently and identically distributed as $A$ and $B$, respectively.

□

Next we show that the sampled Markov chain is positive recurrent, so that a unique stationary distribution exists. The stationary distribution must also satisfy the stochastic recurrence given in the above lemma, which will later lead

us to the regular variation result. But first we need a couple of technical results about the random variables $A$ and $B$.

**Lemma 2.** *Let $A$ and $B$ be the random variables in the previous lemma. There exists $0 < \mu_0 < 1$ such that for all values of $0 < \mu < \mu_0$, if $\lambda > (1-\mu)$, $0 < \alpha < 2\lambda(1-\mu)$ and $E\left(\frac{e^{F(1+\mu)}+e^{F(1-\mu)}}{2}\right)^{\alpha} \neq 2$, then the following hold.*

*(i)* $E \log A < 0$

*(ii)* $EA^{\alpha} = 1$,

*(iii)* $E|B|^{\alpha} < \infty$

*(iv)* $E(A^{\alpha} \log A) < \infty$

*(v)* $P(Ax + B = x) < 1$ *for all $x$.*

*Proof.* (i) We can simplify the expression for $E \log A$ by writing

$$J_1 = E\left(\log \frac{e^{F(1+\mu)} + e^{F(1-\mu)}}{2}\right), \quad J_2 = E\left(\log \frac{e^{F(1+\mu)} - e^{F(1-\mu)}}{2}\right),$$

for $F$ an exponential$(2\lambda)$ random variable. Note that $K = n$ with probability $2^{-n}$, so that

$$\begin{aligned}
E \log A &= E\left[E \log A | K\right] \\
&= \frac{1}{2} E \log\left(\frac{e^{F_1(1+\mu)} + e^{F_1(1-\mu)}}{2}\right) \\
&\quad + \sum_{n=1}^{\infty} \frac{1}{2^{n+1}} E \log\left(\frac{e^{F_1(1+\mu)} - e^{F_1(1-\mu)}}{2} \cdot \frac{e^{F_{K+1}(1+\mu)} - e^{F_{K+1}(1-\mu)}}{2} \cdot \prod_{j=2}^{n} \frac{e^{F_j(1+\mu)} + e^{F_j(1-\mu)}}{2}\right) \\
&= \frac{1}{2} J_1 + \sum_{n=1}^{\infty} \frac{1}{2^{n+1}} (2J_2 + (n-1)J_1) \\
&= J_1 + J_2.
\end{aligned}$$

Now that we have a simple expression for $E \log A$, we note that as $\mu \to 0$, $J_1 \to \frac{1}{2\lambda}$ and $J_2 \to -\infty$. Hence

$$\lim_{\mu \to 0} E \log A = -\infty. \tag{2.7}$$

Furthermore, we claim that $E \log A$ is strictly increasing with respect to $\mu$. Taking derivatives of $J_1$ and $J_2$, we have for $\mu > 0$,

$$\frac{d}{d\mu} J_1 = E\left[F \frac{e^{F(1+\mu)} - e^{F(1-\mu)}}{e^{F(1+\mu)} + e^{F(1-\mu)}}\right] > 0, \quad \frac{d}{d\mu} J_2 = E\left[F \frac{e^{F(1+\mu)} + e^{F(1-\mu)}}{e^{F(1+\mu)} - e^{F(1-\mu)}}\right] > 0. \tag{2.8}$$

It is clear that $E \log A \to \infty$ as $\mu \to \infty$, so combining this with (2.7) and (2.8), let

$$\mu_0 = \inf\{\mu : E \log A \geq 0\} > 0.$$

Therefore if $0 \leq \mu < \mu_0$, $E \log A < 0$.

(ii) Define $h(\alpha) = EA^\alpha$ for $\alpha \geq 0$. Note that $h(0) = 1$, $h''(\alpha) > 0$, and $h'(\alpha) = E[A^\alpha \log \alpha]$. If $\mu$ is such that $E \log A < 0$, then $h'(0) < 0$. Further, $h'(\alpha) > 0$ for $\alpha > 1$, so that $EA^\alpha$ is eventually increasing in $\alpha$.

We will show that for finite values of $\alpha$, $h(\alpha) < \infty$. Since $h(\alpha)$ is eventually increasing in $\alpha$, there must exist an $\alpha > 0$ such that $h(\alpha) = 1$.

With a similar idea as defining $J_1, J_2$ to simplify $E \log A$ , we define

$$I_1 = E\left(\frac{e^{F(1+\mu)} + e^{F(1-\mu)}}{2}\right)^\alpha, \quad I_2 = E\left(\frac{e^{F(1+\mu)} - e^{F(1-\mu)}}{2}\right)^\alpha,$$

to simplify $EA^\alpha$ to the following expression:

$$EA^\alpha = \frac{1}{2} I_1 + I_2^2 \frac{1}{4 - 2I_1}. \tag{2.9}$$

Factoring the numerator of $I_1$, and using the moment generating function for the exponential distribution, we have

$$I_1 = \left(\frac{e^{2\mu} + 1}{2}\right)^\alpha E\left(e^{\alpha F(1-\mu)}\right) = \left(\frac{e^{2\mu} + 1}{2}\right)^\alpha \frac{2\lambda}{2\lambda - \alpha(1-\mu)} < \infty, \quad \alpha(1-\mu) < 2\lambda. \tag{2.10}$$

So for $\alpha < \frac{2\lambda}{1-\mu}$, we have $I_2 \leq I_1 < \infty$. Also assuming that $\alpha$ is such that $I_1 \neq 2$, we have from (2.9) that $EA^\alpha < \infty$.

Now we need to show that there is an $0 < \alpha < 2\lambda/(1-\mu)$ such that $EA^\alpha \geq 1$. Take

$$\tilde{\alpha} = \frac{\lambda}{1-\mu}. \tag{2.11}$$

Then $\tilde{\alpha}$ satisfies the constraint, and for $\alpha = \tilde{\alpha}$,

$$I_1 \geq E\left(e^{F(1-\mu)\tilde{\alpha}}\right) = \left(\frac{2\lambda}{2\lambda - \tilde{\alpha}(1-\mu)}\right) = 2. \tag{2.12}$$

Then $EA^{\tilde{\alpha}} \geq \frac{I_1}{2} \geq 1$. Hence $EA^{\tilde{\alpha}}$ is finite, but greater than 1, for some $\tilde{\alpha} > 0$. Hence there is an $\alpha > 0$ such that $EA^\alpha = 1$.

(iii) Notice that $B$ can be written as follows, with empty products being equal to 1.

$$\begin{aligned} B = &\left(\frac{e^{F_{K+1}(1+\mu)} + e^{F_{K+1}(1-\mu)}}{2}\right) + \prod_{j=1}^{K+1}\left(\frac{e^{F_j(1+\mu)} + e^{F_j(1-\mu)}}{2}\right) \\ &+ \sum_{i=1}^{K+1}\left[\left(\frac{e^{F_i(1+\mu)} - e^{F_i(1-\mu)}}{2}\right)\prod_{j=i+1}^{K}\left(\frac{e^{F_j(1+\mu)} + e^{F_j(1-\mu)}}{2}\right)\right] \\ &+ \left(\frac{e^{F_1(1+\mu)} - e^{F_1(1-\mu)}}{2}\right)\left(\frac{e^{F_{K+1}(1+\mu)} - e^{F_{K+1}(1-\mu)}}{2}\right)\prod_{j=1}^{K}\left(\frac{e^{F_j(1+\mu)} + e^{F_j(1-\mu)}}{2}\right). \end{aligned}$$

It is enough to show finiteness of

$$E\sum_{i=1}^{K+1}\left[\left(\frac{e^{F_i(1+\mu)} - e^{F_i(1-\mu)}}{2}\right)^\alpha \prod_{j=i+1}^{K}\left(\frac{e^{F_j(1+\mu)} + e^{F_j(1-\mu)}}{2}\right)^\alpha\right]. \tag{2.13}$$

Recalling the definitions of $I_1$ and $I_2$ above and that $K \sim$ *geometric*$(1/2)$,

$$E\sum_{i=1}^{K+1}\left[\left(\frac{e^{F_i(1+\mu)} - e^{F_i(1-\mu)}}{2}\right)^\alpha \prod_{j=i+1}^{K}\left(\frac{e^{F_j(1+\mu)} + e^{F_j(1-\mu)}}{2}\right)^\alpha\right] = \sum_{k=0}^{\infty} 2^{-k}\sum_{i=1}^{k+1} I_2 I_1^{k-i}. \tag{2.14}$$

Some manipulations then yield

$$\begin{aligned} \sum_{k=0}^{\infty} 2^{-k}\sum_{i=1}^{k+1} I_2 I_1^{k-i} &= \sum_{k=0}^{\infty} 2^{-k} I_2 I_1^k \sum_{i=1}^{k+1}\left(\frac{1}{I_1}\right)^i = I_2\sum_{k=0}^{\infty}\left(\frac{I_1}{2}\right)^k \frac{1 - 1/I_1^{k+2}}{1 - 1/I_1} \\ &\leq \frac{I_1 I_2}{I_1 - 1}\sum_{k=0}^{\infty}\left(\frac{I_1}{2}\right)^k \end{aligned} \tag{2.15}$$

where we use the fact that $I_1 \neq 1$. Recall $\alpha$ is such that $EA^\alpha = 1$. Therefore

$$2 = I_1 + I_2^2 \frac{1}{2 - I_1} > I_1. \tag{2.16}$$

So (2.15) implies finiteness of (2.13).

(iv) Note $E[A^\alpha \log A] = E[\frac{\partial}{\partial \alpha} A^\alpha]$. If we can switch derivative and expectation, we have

$$E\left[\frac{\partial}{\partial \alpha} A^\alpha\right] = \frac{\partial}{\partial \alpha} E[A^\alpha] = 0 \tag{2.17}$$

by part (ii). We can see that the interchange is valid by letting $Z = A^{1+\alpha}$ and noticing that $A < Z$ a.s. and $E|Z| = E[A^{1+\alpha}] < \infty$.

(v) As shown below in (2.23), $Ax+B$ is a continuous random variable, so $P(Ax+B = x) = 0$.

□

**Theorem 8.** *The Markov chain $(X_n; n \geq 0)$ is positive Harris recurrent and hence there exists a stationary solution $X$ to the stochastic recurrence (2.4) which satisfies*

$$X \stackrel{d}{=} AX + B$$

*for some random variables $A$ and $B$ independent from $X$.*

*Proof.* To prove positive Harris recurrence, we use a result from [61]. Defining for a real function $V$ the expected one-step difference

$$\Delta V(x) = \int P(x, dy)V(y) - V(x) = EV(X_1) - V(x), \tag{2.18}$$

we know that the Markov chain $\{X_n\}$ is positive Harris recurrent if there exists a function $V : [1, \infty] \to [0, \infty]$, a petite set $C \subset \mathcal{B}(\mathbb{R})$, and a constant $b < \infty$ such that $V$ is finite and bounded on $C$ and

$$\Delta V(x) \leq -1 + b\mathbb{1}_{C(x)}, \qquad x \in [1, \infty). \tag{2.19}$$

Let $x$ be the initial state of the Markov chain. Then after one step, we are at state $X_1 = Ax + B$. Let $\alpha > 0$ be such that $EA^\alpha = 1$. Then $EA^\theta < 1$ for any $0 < \theta < \alpha$. Fix such a $\theta$, under the additional condition that $\theta < 1$. Then let

$$V(x) = x^\theta.$$

From the previous lemma, we know that $EB^\theta < \infty$. Now, calculating the one-step difference of $V$, we have

$$\begin{aligned} \Delta V(x) &= EV(Ax + B) - V(x) = E(Ax + B)^\theta - x^\theta \\ &\leq E((Ax)^\theta + B^\theta) - x^\theta = x^\theta(EA^\theta - 1) + EB^\theta \\ &\leq -1 \end{aligned} \tag{2.20}$$

for $x$ sufficiently large.

Let $c > 1$ be such that the above inequality is satisfied for $x > c$. We define $C = [1, c]$, and note that for $x \in C$,

$$\Delta V(x) \leq c^\theta(EA^\theta - 1) := b, \quad x \in C. \tag{2.21}$$

Note that we have set $b = c^\theta(EA^\theta - 1)$ to be the finite constant in (2.19). Then (2.20) and (2.21) imply (2.19). $V$ is clearly bounded and finite on $C$, so the remaining step is to prove that the set $C$ is petite.

Consider the one step transition $x \mapsto Ax + B$. Let $N \sim Geo(1/2)$ and $F_i \sim Exp(2\lambda)$, $i = 1, 2, \ldots$ all be independent. Then we can write

$$\begin{cases} A = 2^{-1}\left(e^{F_1(1+\mu)} + e^{F_1(1-\mu)}\right) & N = 0 \\ A = 2^{-(N+1)} \prod_{j=1}^{2}\left(e^{F_j(1+\mu)} - e^{F_j(1-\mu)}\right) \cdot \prod_{j=3}^{N+1}\left(e^{F_j(1+\mu)} + e^{F_j(1-\mu)}\right) & N \geq 1. \end{cases} \tag{2.22}$$

$B$ can be written similarly. Let $w_1 = e^{F_1}$, $w_2 = e^{F_2}$. In terms of $w_1$ and $w_2$,

$$A = g_1(w_1, w_2) = (w_1^{1+\mu} - w_1^{1-\mu})(w_2^{1+\mu} - w_2^{1-\mu})CK + (w_1^{1+\mu} + w_1^{1-\mu})(1-K),$$

$$B = g_2(w_1, w_2) = \left[\frac{1}{2}(w_1^{1+\mu} + w_1^{1-\mu}) + (w_1^{1+\mu} - w_1^{1-\mu})D + 2(w_1^{1+\mu} - w_1^{1-\mu})w_2^{1+\mu}C\right]K$$
$$+ (w_1^{1+\mu} - w_1^{1-\mu})(1-K),$$

for appropriate choices of $C$ and $D$, and where $K = \mathbb{1}_{N\geq 1}$.

Both $g_1$ and $g_2$ have continuous partial derivatives at all points. The Jacobian is

$$J(w_1, w_2) = \begin{vmatrix} \frac{\partial g_1}{\partial w_1} & \frac{\partial g_1}{\partial w_2} \\ \frac{\partial g_2}{\partial w_1} & \frac{\partial g_2}{\partial w_2} \end{vmatrix}$$
$$= 2CK(1+\mu)(w_1^{1+\mu} - w_1^{1-\mu})w_2^{\mu}\Big((1-K)((1+\mu)w_1^{\mu} + (1-\mu)w_1^{-\mu}) + CK((1+\mu)w_1^{\mu} - (1-\mu)w_1^{-\mu}$$
$$\left(w_2^{1+\mu} - w_2^{1-\mu}\right)\Big) - CK(w_1^{1+\mu} - w_1^{1-\mu})\left((1+\mu)w_2^{\mu} - (1-\mu)w_2^{-\mu}\right)$$
$$\Big((1-K)((1+\mu)w_1^{\mu} - (1-\mu)w_1^{-\mu}) + K\Big(D((1+\mu)(w_1^{\mu} - (1-\mu)w_1^{-\mu})$$
$$+\frac{1}{2}((1+\mu)(w_1^{\mu} + (1-\mu)w_1^{-\mu}) + 2Cw_2^{1+\mu}((1+\mu)(w_1^{\mu} - (1-\mu)w_1^{-\mu})\Big)\Big),$$

which is continuous and nonzero at $w_1 = w_2 = 2$. Therefore there exists a compact set $S$ containing $(2,2)$ such that $J(w_1, w_2) \neq 0$ for $(w_1, w_2) \in S$. Hence the density function is strictly positive in this set.

Let $f(a,b)$ be the joint pdf of $A, B$. Then starting from $x$, the pdf of the first step, $(Ax + B)$, is

$$h_x(y) = \frac{1}{x}\int_0^{\infty} f\left(\frac{y-b}{x}, b\right)db. \tag{2.23}$$

Fix $y$ such that $(\frac{y-b}{x}, b) \in S$ for some value of $b$ for all $x \in [1, c]$. Then

$$\min_{x\in[1,c]} h_x(y) \equiv \phi(y) > 0, \tag{2.24}$$

since the minimum is achieved on a compact set. Now define measure

$$\Phi(E) = \int_E \phi(y)dy \tag{2.25}$$

for $E \in \mathcal{B}(\mathcal{X})$. This measure is nontrivial since $\phi(y)$ is strictly positive for $y$ in a compact set. Finally, starting from any $x \in [1, c]$, the probability that the first step lands in set $E$ is

$$P(Ax + B \in E) = \int_E h_x(y)dy \geq \int_E \phi(y)dy = \Phi(E). \tag{2.26}$$

This proves that $C$ is a petite set. □

**Corollary 1.** *Let $Y$ be distributed as the stationary distribution of $(X_n; n \geq 0)$. Then $Y$ is regularly varying with index $\alpha > 0$, where $\alpha$ is as in Lemma 2. In particular there exists a constant $C > 0$ such that*

$$P(Y > x) \sim Cx^{-\alpha}, \quad x \to \infty.$$

*Proof.* The result follows from Lemma 2 and Theorem 1.1 in [16] □

The results in this section were proved for the Markov chain $(X_n)$ that resulted from sampling the continuous time process $X_1(t)$ at arrivals of the second Poisson process, $N_2$. However, by symmetry of our model, all results are also true for the Markov chain resulting from sampling $X_2(t)$ at arrivals of $N_1$. Hence we now have an understanding of the two-dimensional process at arrivals of *either* Poisson process. This will be the basis for the next section.

## 2.4 The Joint Process

Consider the state of the system at an arrival of *either* of the Poisson processes, that is, at points of the combined counter $N_1 \cup N_2$. Let $W_1, W_2, \ldots,$ be the arrival times of $N_1 \cup N_2$, so that $(W_{n+1} - W_n)$ are i.i.d. exponentially distributed random variables with rate $2\lambda$.

If we randomly pick a time $W_n$, it has an equal chance of being a reset time of $X_1$ or a reset time of $X_2$. If $W_n$ is a reset time of $X_1$, then $X_2$ at time $W_n$ is randomly distributed as the stationary distribution of $(X_n)$, as in the preceding section. If $W_n$ is instead a reset time of $X_2$, then $X_1$ takes on the stationary distribution.

Hence we can consider the discrete-time system $((X_1(W_n), X_2(W_n)); n \geq 1)$. This system has state space

$$([1, \infty) \times \{1\}) \cup (\{1\} \times [1, \infty)$$

and is a Markov chain by markovity of the sampled process. This two-dimensional Markov chain is ergodic, by symmetry and union of the two one-dimensional processes, with stationary distribution of the form

$$(X_1, X_2)(W_n) = \begin{cases} (Y, 1) & \text{w.p. } \frac{1}{2} \\ (1, Y) & \text{w.p. } \frac{1}{2} \end{cases}, \qquad n = 1, 2, \ldots, \tag{2.27}$$

where $Y$ is distributed according to the stationary distribution of $(X_n)$.

From the structure of our model, given in (2.2), we know that the process is deterministic between arrival times of the Poisson processes, as given in (2.3). Therefore the "randomness" in the system is completely characterized by (2.27), and it is straightforward to determine how the process behaves between resets.

With this understanding of the joint process, we are on our way to proving multivariate regular variation. We begin by proving that the joint process is an irreducible T-process. From there, we prove positive Harris recurrence via recurrence of the one-dimensional sampled process. With these results, we know that a unique stationary distribution exists. Combining (2.3) and (2.27), we formulate an expression for the stationary distribution of the joint process.

Finally, in this stationary regime, we prove regular variation and calculate the tail dependence coefficient.

**Lemma 3.** *The joint process* $((X_1(t), X_2(t)); t \geq 0)$ *is an irreducible T-process.*

*Proof.* We prove using a result from [61]. Their Theorem 4.1 states that, since our process is non-evanescent, it is an irreducible T-process if and only if every compact set is petite.

Hence we consider a general compact set $C \subset [1, \infty) \times [1, \infty)$, and we will look at a "one-step" transition to prove that $C$ is petite. Here, the "step" is from initial state $(x_1, x_2)$ to the state at the first arrival of $N_2$. Let $T$ denote the time of the first arrival of $N_2$. Then the state of interest is $(X_1(T), 1)$.

We observe in the continuous case that $X_1(T)$ can be written as

$$X_1(T) = Ax_1 + Bx_2 + C, \tag{2.28}$$

where $A, B, C$ are random variables depending on $T$ and any arrivals of $N_1$ that might have occurred in the interval $(0, T)$. Similarly to the one-dimensional case, we let $(F_i)_{i\geq 1}$ be iid exponential random variables with parameter $2\lambda$. Then we can write $A$, $B$, and $C$ in terms of $w_1 = e^{F_1}$, $w_2 = e^{F_2}$, and $w_3 = e^{F_3}$:

$$A = g_1(w_1, w_2, w_3) = (1 - K)(w_1^{1+\mu} + w_1^{1-\mu}) + KK_1(w_1^{1+\mu} - w_1^{1-\mu})(w_2^{1+\mu} + w_2^{1-\mu})(w_3^{1+\mu} + w_3^{1-\mu})$$
$$B = g_2(w_1, w_2, w_3) = (1 - K)(w_1^{1+\mu} - w_1^{1-\mu}) + KK_1(w_1^{1+\mu} + w_1^{1-\mu})(w_2^{1+\mu} + w_2^{1-\mu})(w_3^{1+\mu} - w_3^{1-\mu})$$
$$C = g_3(w_1, w_2, w_3) = K\left[(w_3^{1+\mu} + w_3^{1-\mu}) + K_1(w_1^{1+\mu} - w_1^{1-\mu})(w_2^{1+\mu} + w_2^{1-\mu}) + K_1(w_2^{1+\mu} - w_2^{1-\mu}) + K_2\right],$$

where $K, K_1, K_2$ are random variables independent from each other and from $(F_i)_{i\geq 1}$. The Jacobian given by these three functions is nonzero for $w_1 = w_2 = w_3 = 2$, and is thus nonzero for all values in some compact set $S$.

Therefore we can define $f(a,b,c)$ as the joint density of $A, B$, and $C$. The density of the first step, that is, the density of $(X_1(T), 1)$ given initial state $(x_1, x_2)$, is given by

$$h_{x_1,x_2}(y_1,y_2) = \mathbb{1}_{y_2=1} \cdot \frac{1}{x_1} \int_{c=0}^{\infty} \int_{b=0}^{\infty} f\left(\frac{y_1 - c - bx_2}{x_1}, b, c\right) dbdc.$$

Fix $y_1$ such that $\left(\frac{y_1-c-bx_2}{x_1}, b, c\right) \in S$ for some $b, c \geq 1$, all $x_1, c_2 \in C$. Then $h_{x_1,x_2}(y_1, 1) > 0$, so we can take

$$\phi(y_1) \equiv \min_{(x_1,x_2)\in C} h_{x_1,x_2}(y_1, 1) > 0.$$

Now define nontrivial measure

$$\Phi(E) = \int_E \phi(y_1)\mathbb{1}_{y_2=1} dy_1 dy_2$$

for $E \in \mathcal{B}(\mathcal{X})$. Then for $(x_1, x_2) \in C$,

$$P^T(x_1, x_2, E) = \int_E h_{x_1,x_2}(y_1, y_2) dy_1 dy_2 = \int_E h_{x_1,x_2}(y_1, 1)\mathbb{1}_{y_2=1} dy_1 dy_2 \geq \int_E \phi(y_1)\mathbb{1}_{y_2=1} = \Phi(E).$$

Hence $C$ is a petite set and $(X_1(t), X_2(t))$ is an irreducible T-process. □

Next we again refer to [61] to prove positive Harris recurrence, by fixing a closed, petite set, and showing that the expected hitting time of this set is finite.

**Lemma 4.** *The joint process* $((X_1(t), X_2(t)); t \geq 0)$ *is positive Harris recurrent.*

*Proof.* Let $C = [1, 2] \times [1, 2]$ and

$$T_1^{(1)}, T_2^{(1)}, T_3^{(1)}, \ldots \text{ be the arrival times of } N_1,$$

$$T_1^{(2)}, T_2^{(2)}, T_3^{(2)}, \ldots \text{ be the arrival times of } N_2.$$

We will show that the first hitting time of $C$ is finite with probability 1. Let $\tau_C = \inf\{t \geq 0 : (X_1(t), X_2(t)) \in C\}$. For any initial state $(x_1, x_2) \in [1, \infty) \times [1, \infty)$ we want to show

$$P_{x_1,x_2}(\tau_C < \infty) = 1. \tag{2.29}$$

Conditioning on the time of the first reset, we must have either $\tau_C \leq \min\{T_1^{(1)}, T_1^{(2)}\}$ or $\tau_C \geq \min\{T_1^{(1)}, T_1^{(2)}\}$. So either the process enters $C$ before the first reset, or we must prove it enters $C$ in finite time after the first reset. At the time of the first reset, the system looks like $(1, x)$ or $(x, 1)$, $x \geq 1$, depending on which variable was reset. By symmetry, it is thus sufficient to prove (2.29) assuming initial state $(x, 1) \in [1, \infty) \times \{1\}$. Hence we now aim to show

$$P_{x,1}(\tau_C < \infty) = 1, \quad x \geq 1. \tag{2.30}$$

By Theorem 8, the Markov chain $(X_1(T_n^{(2)}); n \geq 1) = (X_n; n \geq 1)$ is positive recurrent. Taking $\hat{C} = [1, 2]$, and $\hat{\tau}$ the first time $(X_n)$ hits $\hat{C}$,

$$\hat{\tau} = \inf\{n \geq 1 : X_n \in \hat{C}\}, \tag{2.31}$$

we then have

$$P_x(\hat{\tau} < \infty) = 1, \quad x \geq 1.$$

Therefore with probability 1, there exists $n < \infty$ such that $(X_n, 1) \in C$. This proves (2.30).

We now have Harris recurrence, so the next step is to prove positive Harris recurrence. According to Theorem 4.4 in [61], positive Harris recurrence follows if

$$\sup_{(x_1,x_2)\in C} E_{x_1,x_2}[\tau_C(\delta)] < \infty \quad \text{for some } \delta > 0, \tag{2.32}$$

where $\tau_C(\delta)$ is the first hitting time on $C$ after $\delta$:

$$\tau_C(\delta) = \inf\{t > \delta : (X_1(t), X_2(t)) \in C\}. \tag{2.33}$$

Keeping $C = [1,2]\times[1,2]$ but this time letting $\hat{C} = [1, 1.5]$, let us consider once again the Markov chain $(X_n; n \geq 1)$. By positive recurrence of $(X_n)$, we know that $\sup_{x\in\hat{C}} E_x\hat{\tau} < \infty$. For future reference, set

$$M := \sup_{x\in\hat{C}} E_x\hat{\tau}. \tag{2.34}$$

To prove (2.32), we will bound $\tau_C(\delta)$ by a hitting time related to the one-dimensional, sampled process. We do this by comparing $\tau_C(\delta)$ to arrival times of the Poisson process $N_2$. Intuitively, if we let $T^\delta$ be the first arrival time of $N_2$ after $\delta$, then $T^\delta \geq \delta$, so that $\tau_C(T^\delta) \geq \tau_C(\delta)$.

Furthermore, if $X_n \in \hat{C}$ for some $n$, then at that moment, the joint process takes values $(X_n, 1) \in C$. So any hitting time of the sampled process comes no sooner than the hitting time of the joint process.

More concretely, fix $\delta > 0$ and let

$$n_\delta = \inf\{n \geq 1 : T_n^{(2)} > \delta\} \tag{2.35}$$

be the index of the first arrival of $N_2$ that occurs after time $\delta$. Then put

$$T^\delta = T_{n_\delta}^{(2)} > \delta \tag{2.36}$$

as the first arrival time of $N_2$ that occurs after time $\delta$. Letting

$$\tau_C(T^\delta) = \inf\{t \geq T^\delta : (X_1(t), X_2(t)) \in C\}, \tag{2.37}$$

comparison with (2.33) shows that $\tau_C(\delta) \leq \tau_C(T^\delta)$, so we also have the inequality in expectation:

$$E_{x_1,x_2}\tau_C(\delta) \leq E_{x_1,x_2}\tau_C(T^\delta). \tag{2.38}$$

Next we compare to hitting times of the sampled process. Let

$$\hat{n}_\delta = \inf\{n \geq n_\delta : X_n \in \hat{C}\} \tag{2.39}$$

be the first index greater than $n_\delta$ where $X_n$ hits $\hat{C}$. Then define

$$\hat{T}^\delta = T^{(2)}_{\hat{n}_\delta} \geq T^\delta$$

to be the first time after $T^\delta$ where $X_n \in \hat{C}$. At time $\hat{T}^\delta$, we have

$$X_1(\hat{T}^\delta) = X_{\hat{n}_\delta} \in \hat{C} = [1, 1.5] \quad \text{and} \quad X_2(\hat{T}^\delta) = 1$$

so that

$$\left(X_1(\hat{T}^\delta), X_2(\hat{T}^\delta)\right) \in C.$$

Therefore

$$\tau_C(T^\delta) \leq \hat{T}^\delta. \tag{2.40}$$

Taking expectations and combining with (2.38), we have

$$E_{x_1,x_2}\tau_C(\delta) \leq E_{x_1,x_2}\hat{T}^\delta. \tag{2.41}$$

Considering this bound, our next step is to prove finiteness of $E_{x_1,x_2}\hat{T}^\delta$. To do so, we expand $E_{x_1,x_2}\hat{T}^\delta$.

To simplify notation, denote

$$h(y_1) = E_{y_1}\hat{\tau}, \quad y_1 \geq 1. \tag{2.42}$$

Note that

$$\begin{aligned}\hat{T}^\delta &= T^\delta + \sum_{j=n_\delta}^{\hat{n}_\delta - 1}(T^{(2)}_{j+1} - T^{(2)}_j) \\ &= T^\delta + \sum_{i=1}^{\hat{n}_\delta - n_\delta} Z_i,\end{aligned} \tag{2.43}$$

where $Z_i \sim \exp(\lambda)$, $i \geq 1$ are iid interarrival times of $N_2$. Now by Wald's identity and conditioning on $X_1(\hat{T}^\delta)$,

$$\begin{aligned} E_{x_1,x_2}\hat{T}^\delta &= E_{x_1,x_2}T^\delta + E_{x_1,x_2}\sum_{i=1}^{\hat{n}_\delta - n_\delta} Z_i \\ &= \delta + E[T^\delta - \delta] + E_{x_1,x_2}[\hat{n}_\delta - n_\delta]E_{x_1,x_2}[Z_i] \end{aligned} \tag{2.44}$$

$$= \delta + \frac{1}{\lambda} + \frac{1}{\lambda}E_{x_1,x_2}\left[E_{X_1(\hat{T}^\delta)}\left[\hat{n}_\delta - n_\delta \,\middle|\, X_1(\hat{T}^\delta)\right]\right].$$

But $\hat{n}_\delta - n_\delta = \hat{\tau}$, since this is the number of timesteps until $(X_n)$ hits $\hat{C}$. So by the strong Markov property,

$$E_{x_1,x_2}\hat{T}^\delta = \delta + \frac{1}{\lambda} + \frac{1}{\lambda}E_{x_1,x_2}h\left(X_1(\hat{T}^\delta)\right). \tag{2.45}$$

Now by (2.41), we have

$$E_{x_1,x_2}\tau_C(\delta) \leq \gamma E_{x_1,x_2}h\left(X_1(\hat{T}^\delta)\right) \tag{2.46}$$

for some $\gamma > 0$.

Note that $E_{x_1,x_2}h\left(X_1(\hat{T}^\delta)\right)$ is nondecreasing in $x_1$ and in $x_2$. Let $a > 0$ be large enough that, starting from $(1,1)$, $X_1(a) \geq 2$ and $X_2(a) \geq 2$. Then for $(x_1, x_2) \in C$ and some $p > 0$,

$$\begin{aligned} E_1\hat{\tau} &\geq E_1\left[\hat{\tau}\mathbb{1}(\text{no resets in } [0,a])\right] \\ &= P(\text{no resets in } [0,a])E_1\left[\hat{\tau}|\text{no resets in } [0,a]\right)] \\ &= pE_{x_1,x_2}h(X_1(\hat{T}^\delta)). \end{aligned} \tag{2.47}$$

Furthermore, each reset makes the system decrease in state. Therefore,

$$\begin{aligned} E_{x_1,x_2}h(X_1(T_1^{(2)})) &\geq E_{x_1,x_2}\left[h(X_1(T_1^{(2)}))\mathbb{1}(T_1^{(2)} > \delta)\right] \\ &\geq E_{x_1,x_2}\left[h(X_1(T^\delta))\mathbb{1}(T_1^{(2)} > \delta)\right] \\ &\geq P(T_1^{(2)} > \delta)E_{x_1,x_2}h(X_1(T^\delta)). \end{aligned} \tag{2.48}$$

From (2.46) and (2.47),

$$E_{x_1,x_2}\tau_C(\delta) \leq \gamma E_{x_1,x_2} h(X_1(\hat{T}^\delta)) \leq \gamma E_1 \hat{\tau}$$

We conclude that

$$\sup_{(x_1,x_2)\in C} E_{x_1,x_2}\tau_C(\delta) \leq \gamma M < \infty.$$

Therefore the two-dimensional Markov process is positive recurrent.

□

Note that positive recurrence implies that the process is bounded in probability on average [61]. Now we can calculate the unique stationary distribution of the process, $\pi$. By Theorem 8.1 of the same reference, we conclude that for any Borel set $B \subset \mathbb{R}^2$,

$$\pi(B) = \lim_{R\to\infty} \frac{1}{R} \int_0^R \mathbb{1}\left((X_1(t), X_2(t)) \in B\right) dt \tag{2.49}$$

almost surely for any initial state.

We will break up the right hand side expression according to the arrival times of the Poisson processes. Recall $(W_n, n \geq 1)$ are the increasing sequence of the reset times (of both types). Clearly, $W_n \to \infty$ as $n \to \infty$.

It was explained previously that the two-dimensional Markov chain $((X_1(W_n), X_2(W_n))$ is positive recurrent with stationary distribution $\nu$ given in (2.27). Suppose the two-dimensional continuous-time process initiates with this same initial distribution $\nu$. Let $W_0 = 0$, and consider the process during each time segment $(W_{n+1}, W_n)$. The amount of time spent in a set $B$ during one such

interval is

$$\int_{W_{n-1}}^{W_n} \mathbb{1}\left((X_1(t), X_2(t)) \in B\right) dt, \quad n = 1, 2, \ldots \tag{2.50}$$

Note that this integral is identically distributed for each $n = 1, 2, \ldots$, and hence (2.50) is a stationary process.

Finally we break up (2.49) into a sum of the steps of this process and use the Law of Large Numbers.

$$\begin{aligned}\pi(B) &= \lim_{n\to\infty} \frac{1}{W_n} \sum_{j=1}^{n} \int_{W_{j-1}}^{W_j} \mathbb{1}\left((X_1(t), X_2(t)) \in B\right) dt \\ &= 2\lambda E \int_0^{W_1} \mathbb{1}\left((X_1(t), X_2(t)) \in B\right) dt. \end{aligned} \tag{2.51}$$

Note that there are only two sources of randomness in the expectation expression. First is the initial state of the process, and second is the endpoint $W_1$. Importantly, however, the process is determinisitic in between these two points. Hence we can delve farther into this expression to find a simple closed form for our solution.

**Theorem 9.** *Let $T \sim exp(2\lambda)$, and given $T = t$, put $U \sim U(0, t)$. In the stationary regime,*

$$(X_1, X_2) \stackrel{d}{=} \begin{cases} (YV + W, \quad YW + V) & w.p.\ \frac{1}{2} \\ (V + YW, \quad W + YV) & w.p.\ \frac{1}{2} \end{cases} \tag{2.52}$$

*where*

$$W = \frac{e^{U(1+\mu)} + e^{U(1-\mu)}}{2}, \qquad V = \frac{e^{U(1+\mu)} - e^{U(1-\mu)}}{2} \tag{2.53}$$

*and $Y$ is a random variable distributed according to the stationary distribution of $(X_n; n \geq 0)$ as in Theorem 8.*

*Proof.* We directly compute the stationary distribution by taking the probability (2.51) and conditioning on the time of the first arrival $W_1$. Recall interarrival

times are exponentially distributed with $2\lambda$.

$$P_{st}\left((X_1(\cdot), X_2(\cdot)) \in B\right) = 2\lambda E \int_0^\infty 2\lambda e^{-2\lambda u} du \int_0^u \mathbb{1}((X_1(t), X_2(t)) \in B) dt.$$

Next condition on the intital state of the process, which is distributed according to (2.27). Call $F_Y$ the distribution of the stationary distribution of $\{X_n\}_{n\geq 0}$.

$$P_{st}\left((X_1(\cdot), X_2(\cdot)) \in B\right) = 4\lambda^2 \int_0^\infty e^{-2\lambda u} du \int_1^\infty F_Y(dy) \int_0^u \frac{1}{2} \cdot \left[\mathbb{1}((yv + w, yw + v) \in B) + \mathbb{1}((v + yw, w + yv) \in B)\right] dt,$$

where $(w, v)$ are the deterministic coefficients of (2.3). That is,

$$w = \frac{e^{t(1+\mu)} + e^{t(1-\mu)}}{2}, \qquad v = \frac{e^{t(1+\mu)} - e^{t(1-\mu)}}{2}. \tag{2.54}$$

This implies that

$$(X_1, X_2) \stackrel{d}{=} \begin{cases} (YV + W, \quad YW + V) & \text{w.p. } \frac{1}{2} \\ (V + YW, \quad W + YV) & \text{w.p. } \frac{1}{2} \end{cases} \tag{2.55}$$

with $Y \sim F_Y$ and random variables $W, V$ defined as follows. Let $T \sim \exp(2\lambda)$. Given $T = t$, let $U \sim U(0, t)$. Then

$$W = \frac{e^{U(1+\mu)} + e^{U(1-\mu)}}{2}, \qquad V = \frac{e^{U(1+\mu)} - e^{U(1-\mu)}}{2}. \tag{2.56}$$

□

**Theorem 10.** *The random variable $(X_1, X_2)$ distributed as (2.52) is regularly varying in the multivariate sense.*

*Proof.* By Corollary 1, $Y$ is regularly varying with exponent $\alpha > 0$. We need to check that check that $EW^{\alpha+\epsilon} < \infty$ and $EV^{\alpha+\epsilon} < \infty$ for some $\epsilon > 0$.

$$
\begin{aligned}
EW^{\alpha+\epsilon} &= \int_0^\infty 2\lambda e^{-2\lambda t} dt \int_0^t \frac{1}{t} \left( \frac{e^{u(1+\mu)} + e^{u(1-\mu)}}{2} \right)^{\alpha+\epsilon} du \\
&\leq 2\lambda \int_0^\infty e^{-2\lambda t} dt \int_0^t \frac{1}{t} (e^{u(1+\mu)})^{\alpha+\epsilon} du \\
&= \frac{2\lambda}{(1+\mu)(\alpha+\epsilon)} \int_0^\infty t^{-1} e^{-2\lambda t} \left( e^{t(1+\mu)(\alpha+\epsilon)} - 1 \right) dt.
\end{aligned}
$$

Letting $s = e^{-t}$ and call $(1+\mu)(\alpha+\epsilon) := c$.

$$
EW^{\alpha+\epsilon} \leq \frac{2\lambda}{c} \int_0^1 \frac{s^{2\lambda-1}}{-\ln s} \left( s^{-c} - 1 \right) ds. \tag{2.57}
$$

To prove finiteness of this integral, we need to check that it is finite as $s \to 1$ and $s \to 0$. Using L'Hôspital's rule, the integrand as $s \to 1$ is

$$
\lim_{s \to 1} \frac{s^{2\lambda-1}}{-\ln s} \left( s^{-(\alpha+\epsilon)(1+\mu)} - 1 \right) = \lim_{s \to 1} \frac{-cs^{-(\alpha+\epsilon)(1+\mu)-1}}{-1/s} = c, \tag{2.58}
$$

where $c$ is positive.

Recall that $\alpha(1+\mu) < 2\lambda$. Therefore there is an $\epsilon > 0$ small enough that $c = (\alpha+\epsilon)(1+\mu) < 2\lambda$. Hence $2\lambda - c > 0$. Now, for $s \to 0$,

$$
\lim_{s \to 0} \frac{s^{2\lambda-c-1} - s^{2\lambda-1}}{-\ln s} = \lim_{s \to 0} -(2\lambda - c - 1)s^{2\lambda-c-1} + (2\lambda - 1)s^{2\lambda-1}, \tag{2.59}
$$

which is finite under integration. Hence $EW^{\alpha+\epsilon} < \infty$. Since $V \leq W$, $EV^{\alpha+\epsilon} < \infty$ as well.

Recalling Definition 3, define spectral measure $\sigma$ as follows. For any Borel set $A$,

$$
\begin{aligned}
\sigma(A) = &\frac{1}{2} E \left[ \|(V,W)\|^\alpha \mathbb{1} \left( \frac{(V,W)}{\|(V,W)\|} \in A \right) \right] \frac{1}{E\|(V,W)\|^\alpha} \\
&+ \frac{1}{2} E \left[ \|(W,V)\|^\alpha \mathbb{1} \left( \frac{(W,V)}{\|(W,V)\|} \in A \right) \right] \frac{1}{E\|(W,V)\|^\alpha}.
\end{aligned} \tag{2.60}
$$

To prove that $(X_1, X_2)$ is regularly varying in the stationary regime, we will show that (1.4) is satisfied for this measure $\sigma$. From (2.52),

$$\begin{aligned}
&\lim_{x\to\infty} \frac{P\left(\|(X_1,X_2)\| > bx, \frac{(X_1,X_2)}{\|(X_1,X_2)\|} \in A\right)}{P(\|(X_1,X_2)\| > x)} \\
&\quad = \lim_{x\to\infty} \left[ \frac{1}{2} \frac{P\left(\|(YV+W, YW+V)\| > bx, \frac{(YV+W,YW+V)}{\|(YV+W,YW+V)\|} \in A\right)}{P(\|(YV+W,YW+V)\| > x)} \right. \\
&\quad \left. + \frac{1}{2} \frac{P\left(\|(YW+V, YV+W)\| > bx, \frac{(YW+V,YV+W)}{\|(YW+V,YV+W)\|} \in A\right)}{P(\|(YW+V,YV+W)\| > x)} \right] \\
&\quad = \lim_{x\to\infty} \left[ \frac{1}{2} \frac{P\left(Y\|(V,W)\| > bx, \frac{(V,W)}{\|(V,W)\|} \in A\right)}{P(Y\|(V,W)\| > x)} \right. \\
&\quad \left. + \frac{1}{2} \frac{P\left(Y\|(W,V)\| > bx, \frac{(W,V)}{\|(W,V)\|} \in A\right)}{P(\|Y(W,V)\| > x)} \right] \\
&\quad = \frac{b^{-\alpha}}{2} E\left[ \|(V,W)\|^\alpha \mathbb{1}\left( \frac{(V,W)}{\|(V,W)\|} \right) \in A \right] \frac{1}{E\|(V,W)\|^\alpha} \\
&\quad + \frac{b^{-\alpha}}{2} E\left[ \|(W,V)\|^\alpha \mathbb{1}\left( \frac{(W,V)}{\|(W,V)\|} \right) \in A \right] \frac{1}{E\|(W,V)\|^\alpha} \\
&\quad = b^{-\alpha}\sigma(A).
\end{aligned}$$

Hence $(X_1, X_2)$ are regularly varying. □

Now we can compute the asymptotic dependence of $X_1$ and $X_2$.

**Theorem 11.** *The tail dependence coefficient is given by*

$$\lim_{t\to\infty} P(X_2 > t | X_1 > t) = \frac{2EV^\alpha}{EV^\alpha + EW^\alpha}. \tag{2.61}$$

*Proof.* For $x_1, x_2 > 0$,

$$\begin{aligned}
\frac{P(X_1 > tx_1, X_2 > tx_2)}{P(Y > t)} &= \frac{1}{2} \frac{P(YV > tx_1, YW > tx_2)}{P(Y > t)} + \frac{1}{2} \frac{P(YW > tx_1, YV > tx_2)}{P(Y > t)} \\
&= \frac{1}{2} \frac{P(Y > t \max\{\frac{x_1}{V}, \frac{x_2}{W}\})}{P(Y > t)} + \frac{1}{2} \frac{P(Y > t \max\{\frac{x_1}{W}, \frac{x_2}{V}\})}{P(Y > t)} \\
&\xrightarrow{t\to\infty} \frac{1}{2} E\left[ \min\left\{ \frac{V}{x_1}, \frac{W}{x_2} \right\} \right]^\alpha + \frac{1}{2} E\left[ \min\left\{ \frac{W}{x_1}, \frac{V}{x_2} \right\} \right]^\alpha .
\end{aligned} \tag{2.62}$$

So the conditional probability is

$$\begin{aligned} P(X_2 > tx_2 | X_1 > tx_1) &= \frac{P(X_1 > tx_1, X_2 > tx_2)}{P(X_1 > tx_1)} \\ &= \frac{P(X_1 > tx_1, X_2 > tx_2)}{P(Y > t)} \frac{P(Y > t)}{P(X_1 > tx_1)} \\ &\xrightarrow{t\to\infty} \frac{E\left[\min\left\{\frac{V}{x_1}, \frac{W}{x_2}\right\}\right]^\alpha + E\left[\min\left\{\frac{W}{x_1}, \frac{V}{x_2}\right\}\right]^\alpha}{E\left(\frac{V}{x_1}\right)^\alpha + E\left(\frac{W}{x_2}\right)^\alpha}. \end{aligned} \tag{2.63}$$

Taking $x_1 = x_2$ proves the theorem. □

## 2.5 Conclusion

We have shown that the symmetric two-dimensions Poisson counter driven stochastic differential equation has a stationary solution $(X_1, X_2)$. Further, the joint distribution of $(X_1, X_2)$ is regularly varying with positive tail dependence coefficient.

Though we have only proved the symmetric case, we are confident that these proof methods may be extended to the asymmetric case. Similar techniques may also be used in greater dimensions.

CHAPTER 3

# COMMUNITY-AWARE PREFERENTIAL ATTACHMENT MODEL

Social networks often exhibit community structure, where users belong to relatively dense groups called communities. We introduce a community-based, growing random graph model where edges are proposed according to a preferential attachment mechanism on node in-degree, and then formed with an acceptance probability that depends on the community membership of the two end nodes. Considering in-degree as a random vector with components corresponding to in-degree from each community, we show that the asymptotic joint distribution is regularly varying with a measure that concentrates on the line representing the community acceptance probabilities.

## 3.1 Introduction

Many real-life networks are known to exhibit a *scale-free nature* and a *community structure*. A graph is said to have a scale-free nature if it has a power-law type degree distribution asymptotically. A graph is said to exhibit a community structure if it tends to form groups of nodes with a relatively high density of edges between themselves. Examples of such networks include citation graphs, online social networks [27, 6, 51], instant messaging networks [81], and protein interactions [66] among others [64]. There are well known theoretical models that exhibit such phenomena in isolation. In this chapter we develop a community aware preferential attachment model and study the effect of the community on the degree distribution of the nodes.

A graph is said to have a power law degree distribution if for large values of

$k$, the fraction of nodes with degree higher than $k$ is proportional to $k^{-\tau}$ for some constant $\tau > 0$, [27]. The power law degree distribution is a natural phenomenon in many real-life networks. For example, in the citation graph, researchers are more likely to cite the famous papers. Similarly, in a social network people are more inclined to follow the famous people in the network.

The *Preferential Attachment Model* (PAM), proposed in [8], is a commonly used graph generation process that is known to be scale-free, see [14]. In a preferential attachment model, nodes are added to the graph as time progresses and new nodes form edges with other nodes with probability proportional to the node's degree. A node of higher degree is more likely to increase its degree than a node with lower degree and this results in a power-law type heavy-tailed degree distribution [39].

There are many variants of the PAM, such as models with nonlinear preferential attachment and models with a general fitness function [39]. Directed models may differentiate between in-and out-degree of nodes, as in [12], which proves distinct power laws for the in- and out-degree distributions. By considering in- and out-degree simultaneously, [79] go further and prove nonstandard multivariate regular variation for the joint degree distribution.

Communities also appear naturally in many real-life networks. Online social network reflect the communities that users naturally belong to in their life, for example work-place and school. Network of academic collaborations or citation graphs reflect the communities around common research interests of authors. There is however no standard technical definition of a community [31]. The community structure and detection literature can be broadly divided into two topics, that of overlapping and non-overlapping communities. In the non-

overlapping case, each node in the graph is a member of only one community. In the overlapping case, a node can be involved in multiple communities [66]. In this chapter, we assume non-overlapping communities, and we define a communities to be a group of nodes such that an edge between two nodes exists with a probability that depends on the community membership of the nodes.

A *Stochastic Block Model* (SBM) is a common generative graph model in which nodes are divided into blocks, or communities. The SBM is defined by three parameters: the number of nodes $n$, a probability vector $p = (p_1, \ldots, p_k)$, and a symmetric $k \times k$ matrix $Q$ with entries in $[0, 1]$. Each node $i$ is assigned a community $C_i \in \{1, \ldots, k\}$ with probability $p_{C_i}$, and nodes $i$ and $j$ are connected with probability $Q_{C_i C_j}$ independently of all other pairs of nodes [40]. If $Q$ is a matrix with large values in the diagonal and relatively small values in off-diagonal, then we will have a graph with densely connected groups (blocks or communities) of nodes and relatively fewer connections between the groups [47]. Community detection on the SBM is well-studied, and it has been shown that under certain conditions on the edge probabilities, exact recovery is possible [25], [1].

The SBM has been fit to a citation network in [43] and a variant of SBM has been applied to a metabolism network in [24]. However, there is a limit in the applicability of the classical SBM because the degree distribution does not match most real networks. One proposed solution is a degree corrected SBM that introduces further parameters in order to induce a power law degree distribution [47, 74]. Another solution is to consider a special case of the geometric preferential attachment model in [46] which incorporates community membership in a multiplicative fitness function. A community detection algorithm for this model is derived and evaluated in [38]. Another approach taken in [41, 90]

creates community structure by alternating preferential attachment steps and triangle-closing steps, where edges are formed between nodes that have edges with common nodes. In [93] the authors evaluated the consistency of community detection under such a model.

In this chapter we propose a community aware preferential attachment model that incorporates a community structure and characterize the degree distribution of such a graph. Our model is unique in that the characteristics of community structure and power law degree distribution arise due to the division of edge formation into two steps: invitation and response. This is a realistic edge formation process found in online social networks such as Facebook and LinkedIn, where a user sends an invitation to another user, and then the second user either accepts or rejects this invitation. In our model, an invitation is sent to a node with probability proportional its in-degree, and the invitation is accepted with a probability depending on the community membership of both nodes. The edge is formed only if the invitation is accepted.

Another unique contribution is that we analyze the joint distribution of the degree vector denoting the number of edges from a node to the different communities. We show that the distribution is asymptotically jointly regular varying via a novel coupling proof. The proof involves coupling with a graph in which we fix the origin community and time of each edge and specify a fixed proportion of edges to terminate in each community. The asymptotic equivalence of the degree distributions has implications for other models and real-world networks that employ the invitation and response steps.

The chapter is organized as follows. We describe the model and state the main results in section 3.2. We give the proofs for the special case of two sym-

metric communities in sections 3.3 and 3.4. We then discuss the general case in section 3.5.

## 3.2 Model

We develop a generative directed graph model which we call *community aware preferential attachment model* or CAPAM. A CAPAM has $k$ non-overlapping communities and edges in the graph are formed in a 2-step process- first where a node sends an invitation to another node to form an edge and second, the recipient node decides to accept or decline the invitation. An edge is formed only when the recipient node accept the sender's invitation. In this model nodes are added to the graph and they send edge invitations following the usual preferential attachment mechanism [8]. The probability of accepting an invitation depends upon the community membership of the sender and recipient- a node in Community $j$ accepts an invitation from a node in Community $i$ with probability $q_{ij}$.

It is important to setup some notations and a proper definition of the model. The CAPAM denotes a growing sequence of directed graphs, $G = (G(0), G(1), G(2), \ldots)$. Set $G(0)$ to be the graph with $k$ nodes and zero edges with one node in each of the $k$ communities. We grow the graph from $G(t)$ to $G(t+1)$ with the following mechanism:

- with probability $\alpha$, a new node is added to $G(t)$ and its community membership is chosen uniformly from $\{1, \ldots, k\}$. We assign $V_t$ to be this new node.

- with probability $1 - \alpha$, we assign $V_t$ to be a random node from $G(t)$.
- $V_t$ sends an edge invitation to $W_t$, chosen with probability

$$P(W_t = w) = \frac{D_t^w + \delta}{m(t) + \delta(n(t) + k)},$$

  where $D_t^w$ is the in-degree of node $w$ at time $t$, $m(t)$ is the total number of edges in $G(t)$, and $n(t)$ is the number of nodes that have arrived between times 0 and $t$, so that $n(t) + k$ is the total number of nodes in $G(t)$.
- with probability $q_{V_t,W_t}$, $W_t$ accepts the invitation and the directed edge $(V_t, W_t)$ is formed.

In the case that $k = 1$, all nodes belong to the same community, and our model reduces to a variation of PA where each proposed edge is formed or not according to a fixed probability. This model is equivalent to a formulation in which a random number, $W \in \{0, 1\}$, of edges are added at each step $t$. This is a special case of the models described in [23] and [3], who allow a random number $W$ of edges to be added at each time step, where $W$ follows a general probability distribution.

In the presence of communities, the in-degree of a node $v$ in $G(t)$ is a $k$ dimensional vector $\mathbf{D}_t^v$

$$\mathbf{D}_t^v := \left(D_{t,1}^v, D_{t,2}^v, \ldots, D_{t,k}^v\right),$$

where $D_{t,i}^v$ is the number of edges directed to node $v$ from community $i$ in $G(t)$. Our main result establishes the asymptotic properties of the joint distribution of in-degree as time $t \to \infty$. Furthermore, for any $\mathbf{d} = (d_1, d_2, \ldots, d_k)$ let $N_{t,i}(\mathbf{d})$ be the number of nodes in community $i$ with in-degree vector equal to $\mathbf{d}$ at time $t$,

$$N_{t,i}(\mathbf{d}) = \sum_{v=1}^{n_i(t)} \mathbb{1}(\mathbf{D}_t^v = \mathbf{d}),$$

where $n_i(t)$ is the number of nodes in $G(t)$ belonging to community $i$. By definition, $N_{t,i}(\mathbf{d}) = 0$ for any $\mathbf{d} \not\geq 0$. Similarly, let $P_{t,i}(\mathbf{d})$ denote the proportion of nodes in community $i$ with degree vector equal to $\mathbf{d}$ at time $t$

$$P_{t,i}(\mathbf{d}) := \frac{N_{t,i}(\mathbf{d})}{n_i(t)}.$$

Before stating the results we make the following simplifying assumption for the model. This is mainly to reduce complexity that does not add much insight on the problem.

**Assumption 1.** *Suppose that the acceptance probabilities $q_{ij} \in (0,1)$, $i, j = 1, \ldots, k$ satisfy*

$$\frac{1}{k}\sum_{i=1}^{k} q_{ij} = q, \qquad i = 1, \ldots, k.$$

This implies that the average invitation acceptance probability is identical for all communities. Under this assumption, the expected total number of edges in $G(t)$ is $qt$ at any time $t$.

We will see that as $t \to \infty$, $P_{t,i}(\cdot)$ converges to a probability distribution $p_i(\cdot)$ that is defined by the recursion

$$p_i(\mathbf{d}) = \mathbb{1}_{\mathbf{d}=\mathbf{0}} - \frac{(\sum_j d_j + \delta)(\sum_j q_{ji})}{k(q+\alpha\delta)} p_i(\mathbf{d}) + \frac{\sum_j d_j - 1 + \delta}{k(q+\alpha\delta)} \sum_j q_{ji} p_i(\mathbf{d} - \mathbf{e}_j), \tag{3.1}$$

where $\mathbf{e}_j$ is the $j$th unit vector with 1 in the $j$th coordinate and zeros elsewhere. The solution to (3.1) is

$$p_i(\mathbf{d}) = (c+1)\binom{\sum_j d_j}{d_1, d_2, \ldots, d_k}\left(\prod_{j=1}^{k}\left(\frac{q_{ji}}{kq}\right)^{d_j}\right)\frac{\Gamma(\sum_j d_j + \delta)\Gamma(1+c+\delta)}{\Gamma(\sum_j d_j + 2 + c + \delta)\Gamma(\delta)}, \quad \mathbf{d} \geq 0 \tag{3.2}$$

where $c = \frac{\alpha\delta}{q}$ and $\binom{\sum_j d_j}{d_1, d_2, \ldots, d_k}$ is the multinomial coefficient.

As such, a node belonging to community $i$ of the asymptotic graph can be said to have a random degree vector with distribution $p_i(\cdot)$. Our main results

state that the degree distribution of nodes belonging to community $i$ in $G(t)$ converges to $p_i(\cdot)$, and the asymptotic distribution is multivariate regularly varying with index $1 + c$.

**Theorem 12.** *For each $i = 1, \ldots, k$, there exists a constant $C > 0$ such that, as $t \to \infty$,*

$$P\left(\max_{\mathbf{d} \geq 0} |P_{t,i}(\mathbf{d}) - p_i(\mathbf{d})| \geq Ct^{-2/5}\right) = o(1).$$

**Theorem 13.** *Let $\mathbf{D_i}$ be a random vector from distribution $p_i(\cdot)$, $i = 1, \ldots, k$. Let $\nu_i(\cdot)$ be the measure concentrating on the ray $\{(q_{1i}, \ldots, q_{ki})t : t > 0\}$ such that the set*

$$A = \left\{ \mathbf{d} \in \mathbb{R}^k : \|\mathbf{d}\| > a, \frac{\mathbf{d}}{\|\mathbf{d}\|} = \frac{(q_{1i}, \ldots, q_{ki})}{\sqrt{q_{1i}^2 + \cdots + q_{ki}^2}} \right\}$$

*has measure*

$$\nu_i(A) = c_i a^{-(c+1)},$$

*where $c_i$ is given in (3.25). Then as $x \to \infty$,*

$$x^{c+1} P\left(\frac{1}{x}\mathbf{D} \in \cdot\right) \xrightarrow{v} \nu_i(\cdot).$$

As mentioned in Section 3.1, these results hold for any $k \geq 1$. The special case where $k = 1$ has an alternate interpretation [3], [23]. The results for $k \geq 2$ are new to our knowledge.

For clarity, the majority of this chapter focuses on the particular case where $k = 2$ and the acceptance probabilities are symmetric. Theorems 12 and 13 are restated and proved for this case in Sections 3.3 and 3.4, respectively. For a brief discussion of the general case, see Section 3.5.

## 3.3 CAPAM with 2 Symmetric Communities

It is insightful to go through the proof of Theorem 12 for a CAPAM with 2 communities, which is the focus of this section. We assume that the communities are symmetric, i.e., $q_{11} = q_{22} = q_{with}$ and $q_{12} = q_{21} = q_{bet}$. In this formulation, the probability that a recipient will accept an edge invitation depends only on whether the sender and the recipient are in the same community or not. We restate and prove Theorem 12 for this case. Because of the symmetry, it is enough to prove the results for community 1 without any loss of generality.

Let $N_t(i, j)$ denote the number of nodes in community 1 with $i$ edges originating in community 1, and $j$ edges originating in community 2. We follow the convention that $N_t(i, j) = 0$ if $i < 0$ or $j < 0$. Define the proportion of nodes in community 1 with in-degree $(i, j)$ as $P_t(i, j) = \frac{N_t(i,j)}{n_1(t)}$.

For the sake of reducing complexity in the proof method, we make the following assumptions.

**Assumption 2.** *Suppose* $q = \frac{1}{2}(q_{with} + q_{bet}) \in (0, 1)$ *is rational,* $\alpha \in (0, 1)$ *is rational, and*

$$q_{with} < q + \alpha\delta, \qquad q_{bet} < q + \alpha\delta. \tag{3.3}$$

The conditions in (3.3) can be interpreted as $\delta$ and the node arrival rate $\alpha$ being large enough to compensate for disparity between $q_{with}$ and $q_{bet}$. Although the forthcoming proof relies on these assumptions, the above conditions may not be necessary for the result.

Theorem 12 gives the asymptotic distribution on a node's degree, conditional upon its community membership. In the setting of two symmetric com-

munities, we will state and prove our results with respect to nodes belonging to community 1, without loss of generality. Let $N_t(i, j)$ be the number of nodes belonging to community 1 with in-degree vector equal to $(i, j)$ at time $t$¿ Similarly, let $P_t(i, j)$ be the proportion of community 1 nodes with degree $(i, j)$. Then Theorem 12 implies that $P_t(i, j)$ converges in probability to $p(i, j)$, where $p(i, j)$ satisfies the recursion

$$p(i,j) = \mathbb{1}_{i=j=0} - \frac{i+j+\delta}{1+c}p(i,j) + \frac{i+j-1+\delta}{1+c}\left[\frac{q_{with}}{2q}p(i-1,j) + \frac{q_{bet}}{2q}p(i,j-1)\right], \quad (3.4)$$

for $i, j \geq 0$. By convention, $p(i, j) = 0$ for $i < 0$ or $j < 0$, and therefore (3.4) solves to

$$p(i,j) = (1+c)\binom{i+j}{i}\left(\frac{q_{with}}{2q}\right)^i\left(\frac{q_{bet}}{2q}\right)^j\frac{\Gamma(i+j+\delta)\Gamma(1+c+\delta)}{\Gamma(i+j+2+c+\delta)\Gamma(\delta)}, \qquad i,j \geq 0. \quad (3.5)$$

The result that we will prove in this section is as follows.

**Theorem 14.** *There exists a constant $C > 0$ such that*

$$P\left(\max_{i,j\geq 0}|P_t(i,j) - p(i,j)| \geq Ct^{-2/5}\right) = o(1)$$

*as $t \to \infty$.*

Central to our proof of Theorem 14 is a coupling between the CAPAM graph, $G(t)$ and a graph $\hat{G}(t)$ that has similar generative rules but a deterministic number of nodes and edges. At any time $t \geq 0$, $\hat{G}(t)$ has exactly $\lfloor \alpha t \rfloor + 2$ nodes and $\lfloor qt \rfloor$ edges. The initial graph $\hat{G}(0)$ is defined identically to $G(0)$, and for $t \geq 0$, $\hat{G}(t+1)$ and evolves from $\hat{G}(t)$ as follows:

- Set $i = 1$ if $\lfloor \alpha(t+1) \rfloor$ is odd, and $i = 2$ if $\lfloor \alpha(t+1) \rfloor$ is even
- If $\lfloor \alpha(t+1) \rfloor > \lfloor \alpha t \rfloor$, a new node $\hat{V}_t$ is added to the graph and assigned membership to community $i$. Otherwise, we assign $\hat{V}_t$ to a node randomly selected from community $i$.

- If $\lfloor q(t+1) \rfloor > \lfloor qt \rfloor$

  - Assign a value to $J \in \{1, 2\}$ according to

$$P(J = j) = \begin{cases} \frac{q_{with}}{2q} & j = i \\ \frac{q_{bet}}{2q} & j \neq i \end{cases}.$$

  - Given $J$, assign $\hat{W}_t$ according to

$$P(\hat{W}_t = w) = \frac{\hat{D}_t^w + \delta}{\sum_{u \in C_J}(\hat{D}_t^u + \delta)},$$

    and form the directed edge $(\hat{V}_t, \hat{W}_t)$.

Analogous to the definition of $N_t(i, j)$, define $\hat{N}_t(i, j)$ be the random number of vertices in community 1 of $\hat{G}(t)$ with in-degree $(i, j)$,

$$\hat{N}_t(i, j) = \sum_{v \in \hat{C}_1(t)} \mathbb{1}(\hat{D}_t^v = (i, j)).$$

The proof of Theorem 14 is essentially broken down into three steps. First, we show that $G(t)$ and $\hat{G}(t)$ have similar degree distributions by comparing $N_t(i, j)$ with $\hat{N}_t(i, j)$. Second, we show that in $\hat{G}$, the degree distribution concentrates around its expectation. Third, we compare the expected degree distribution of $\hat{G}$ to the $p(i, j)$ distribution given in (3.4). These steps translate into the three lemmas stated below.

**Lemma 5.** *Let*

$$t^* = \sup\{t \in \mathbb{Z}^+ : q - \frac{1+\alpha}{2\alpha}\sqrt{\frac{\log t}{t}} \leq 0\},$$

$$\eta = \min\left\{\frac{2\delta}{t*},\ q - \frac{1+\alpha}{2\alpha}\sqrt{\frac{\log(t^*+1)}{t^*+1}} + 2\delta\right\},$$

*and*

$$C_X = 3q + \frac{q}{\eta}\frac{1+\alpha}{2\alpha} + \frac{\delta q + \left(\delta + \frac{3}{2}\right)(1-\alpha)|q_{with} - q_{bet}|}{2\sqrt{\alpha}}. \tag{3.6}$$

*Then for any $C > 4(C_X + 1)$,*

$$P\left(\max_{i,j} \left|N_t(i,j) - \hat{N}_t(i,j)\right| \geq Ct^{3/5}\right) = o(1)$$

*as $t \to \infty$.*

**Lemma 6.** *For any $C > 4$,*

$$P\left(\max_{i,j\geq 0} \left|\hat{N}_t(i,j) - E[\hat{N}_t(i,j)]\right| \geq Ct^{3/5}\right) = o(1)$$

*as $t \to \infty$.*

**Lemma 7.** *Let*

$$K = \min\left\{k \geq 0 : kT \geq \max\left\{4,\ \frac{\delta}{q}\left(\frac{q_{with}}{q + \alpha\delta - q_{with}}\right),\ \frac{\delta}{q}\left(\frac{q_{bet}}{q + \alpha\delta - q_{bet}}\right)\right\}\right\}$$

*and*

$$C_\epsilon = \max\left\{T\left(1 + \frac{\alpha}{2} + \frac{3\alpha}{q}\right),\ KT,\ \left(\frac{2q}{q_{with}} + \frac{2q}{q_{bet}}\right)\left(\frac{T+\delta}{(KT)^{3/5} - 2}\right)\right\}. \tag{3.7}$$

*Then for all $t \geq 0$,*

$$\max_{i,j\geq 0}\left|E[\hat{N}_t(i,j)] - \frac{\alpha}{2}tp(i,j)\right| \leq C_\epsilon(t^{3/5} + 1).$$

The three lemmas are proved in Sections 3.3.1, 3.3.2, and 3.3.3, respectively. But first, we combine their results to prove Theorem 14.

*Proof of Theorem 14.* Suppose $C \geq \max\{4,\ C_\epsilon,\ 4(C_X + 1)\}$. For any $i, j \geq 0$,

$$|P_t(i,j) - p(i,j)| \leq |N_t(i,j)|\left|\frac{1}{n_1(t)} - \frac{1}{\frac{\alpha}{2}t}\right| + \left|\frac{N_t(i,j)}{\frac{\alpha}{2}t} - p(i,j)\right|.$$

Since $|N_t(i,j)| \leq n_1(t)$ by definition,

$$|N_t(i,j)|\left|\frac{1}{n_1(t)} - \frac{1}{\frac{\alpha}{2}t}\right| = \frac{|\frac{\alpha}{2}t - n_1(t)|}{\frac{\alpha}{2}t}.$$

Therefore

$$P\left(\max_{i,j\geq 0}|P_t(i,j)-p(i,j)|\geq \frac{12}{\alpha}Ct^{-2/5}\right)$$
$$\leq P\left(\left|\frac{\alpha}{2}t-n_1(t)\right|\geq 3Ct^{3/5}-1\right)+P\left(\max_{i,j\geq 0}\left|N_t(i,j)-\frac{\alpha}{2}p(i,j)\right|\geq 3Ct^{3/5}+1\right).$$

Since $n_1(t)$ is one more than a binomial random variable on $t$ trials with success probability $\frac{\alpha}{2}$, Hoeffding's inequality implies

$$P\left(\left|\frac{\alpha}{2}t-n_1(t)\right|\geq 3Ct^{3/5}-1\right)\leq 2\exp\left(-2(3Ct^{3/5}-2)^2/t\right)=o(1).$$

By Lemmas 5, 6, and 7,

$$P\left(\max_{i,j\geq 0}\left|N_t(i,j)-\frac{\alpha}{2}tp(i,j)\right|\geq C(3t^{3/5}+1)\right)$$
$$\leq P\left(\max_{i,j\geq 0}\left|N_t(i,j)-\hat{N}_t(i,j)]\right|\geq Ct^{3/5})\right)+P\left(\max_{i,j\geq 0}\left|\hat{N}_t(i,j)-E[\hat{N}_t(i,j)]\right|\geq Ct^{3/5})\right)$$
$$+P\left(\max_{i,j\geq 0}\left|E[\hat{N}_t(i,j)]-\frac{\alpha}{2}tp(i,j)\right|\geq C(t^{3/5}+1)\right)$$
$$=o(1).$$

□

Before moving on to the proofs of Lemmas 5, 6, and 7, we take a moment to define a quantity $T$ that will be used in proving Lemmas 5 and 7.

Note that the time between node arrivals in $\hat{G}$ is either $\lfloor 1/\alpha\rfloor$ or $\lceil 1/\alpha\rceil$, and similarly, edge interarrival times are either $\lfloor 1/q\rfloor$ or $\lceil 1/q\rceil$. Because $\alpha$ and $q$ are rational, the sequences of node and edge interarrival times are periodic. Further, when community membership $i$ of $\hat{V}_t$ is considered, there is another periodic pattern describing the interarrival times of nodes to community 1, for instance. Because of this periodic nature, it will be useful to define a period $T$ that demarcates all of these patterns in combination.

Define the indicator functions

$$\mathbb{1}_{\alpha,t} = \mathbb{1}(\lfloor \alpha(t+1) \rfloor > \lfloor \alpha t \rfloor)$$

$$\mathbb{1}_{q,t} = \mathbb{1}(\lfloor q(t+1) \rfloor > \lfloor q t \rfloor)$$

$$\mathbb{1}_{\text{odd},t} = \mathbb{1}(\lfloor \alpha(t+1) \rfloor \text{ is odd}).$$

Let the period $T$ be such that for all $t \geq 0$,

$$\mathbb{1}_{q,t} = \mathbb{1}_{q,t \bmod T}$$

$$\mathbb{1}_{\alpha,t} = \mathbb{1}_{\alpha,t \bmod T}$$

$$\mathbb{1}_{\text{odd},t} = \mathbb{1}_{\text{odd},t \bmod T}$$

and the following identities hold: for any nonnegative integer $k$,

$$\sum_{t=kT}^{kT+T-1} \mathbb{1}_{q,t} = \sum_{t=0}^{T-1} \mathbb{1}_{q,t} = qT, \tag{3.8}$$

and

$$\sum_{t=0}^{T-1} \mathbb{1}_{q,t}\mathbb{1}_{\text{odd},t} = \sum_{t=0}^{T-1} \mathbb{1}_{q,t}(1 - \mathbb{1}_{\text{odd},t}) = \frac{qT}{2} \tag{3.9}$$

$$\sum_{t=0}^{T-1} \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t} = \sum_{t=0}^{T-1} \mathbb{1}_{\alpha,t}(1 - \mathbb{1}_{\text{odd},t}) = \frac{\alpha T}{2}. \tag{3.10}$$

The period $T$ will be used in particular to consider the number of edges in $\hat{G}(t)$ that originate from a node in community 1. The defintion of $T$ indicates that, on average, an edge originates in community 1 every $\frac{2}{q}$ timesteps. However, the actual number of community 1 edges may deviate from $\frac{qt}{2}$ by at most a fixed amount.

### 3.3.1 Coupling $N_t(i, j)$ and $\hat{N}_t(i, j)$

Both $G$ and $\hat{G}$ are designed to have about $\alpha t$ nodes and $qt$ edges at time $t$ and similar marginal distributions for the terminal nodes $W_t$ and $\hat{W}_t$. Thus we expect

the asymptotic in-degree distributions to be similar between $G$ and $\hat{G}$, as stated in Lemma 5.

To prove Lemma 5, we devise a coupling scheme to compare the degree distributions of $G(t)$ and $\hat{G}(t)$ at any time $t$. The aim is to select the same terminal nodes $W_t$ and $\hat{W}_t$ at each time step, so that each node in $G(t)$ has the same in-degree as the corresponding node in $\hat{G}(t)$. Complicating matters, $G$ and $\hat{G}$ may not have the same number of nodes nor the same number of edges any given time point. This must be taken into account when we compare $N_t(i, j)$ and $\hat{N}_t(i, j)$ , and it also means that the marginal distributions for $W_t$ and $\hat{W}_t$ do not have the same denominator.

We will deal with these issues by first noting that, although $G(t)$ and $\hat{G}(t)$ will not have exactly the same number of nodes, their node set will be similar, and the number of "excess" nodes remains small relative to total graph size. With regards to edges, we work around the issue by consider the arrival times of edges in the two graphs. Then, instead of comparing in-degree at time $t$, we will compare in-degree at the arrival times of the $k$th edge, even though these arrival times may differ between $G$ and $\hat{G}$. Effectively, this is changing the graph indexing, so that we index $G$ and $\hat{G}$ by number of edges $k$ instead of by elapsed time $t$. Although the arrival time $\tau_k$ of the $k$th edge in $G$ may digress from the arrival time $\hat{\tau}_k$ of the $k$th edge in $\hat{G}$, we will show that the difference in arrival times $|\tau_k - \hat{\tau}_k|$ remains small.

As we are giving results for community 1 nodes without loss of generality, we need only consider the arrival times of edges terminating in community 1.

The result that $G$ and $\hat{G}$ have similar in-degree distributions shows a sur-

prising result in the order of edge operations. In $G$, a potential terminal node $W_t$ is selected, and then the edge may or may not appear, depending on the community membership of $W_t$. However, in $\hat{G}$, it is first decided whether or not to add a new edge, and to which community, and then the terminal node $\hat{W}_t$ is selected *from the specified community*. These processes are antithetical in terms of the dependence on community: one establishes the terminal community before identifying a particular node, and the other determines the potential terminal node without regard to community. However, the fascinating result of the coupling is that, in the long run, both methods result in the same in-degree distribution.

We detail the coupling scheme and prove Lemma 5 in Appendix A, as they are quite involved.

### 3.3.2 Concentration of $\hat{N}_t(i, j)$ around the mean

The objective in coupling $N_t(i, j)$ and $\hat{N}_t(i, j)$, as in Lemma 5, is that we can now study the simpler graph $\hat{G}$ in order to gain insight about the in-degree distribution of $G$. Now that we have shown that $G(t)$ and $\hat{G}(t)$ have similar in-degree distributions, we can proceed by finding the in-degree distribution of $\hat{G}(t)$. The first step of this process is to show that $\hat{N}_t(i, j)$ does not stray far from its mean, $E[\hat{N}_t(i, j)]$. Afterwards, in Section 3.3.3, we will give the limiting distribution of the mean.

To prove Lemma 6, we will define a martingale related to $\hat{N}_t(i, j)$, and then employ the Azuma-Hoeffding inequality to prove concentration around the mean.

Fix $i$, $j$ and define, for $n \in \{0, \ldots, t\}$,

$$M_n = E\left[\hat{N}_t(i,j) \mid \mathcal{F}_n\right],$$

where $\mathcal{F}_n = \sigma\left(\hat{G}(0), \hat{G}(1), \ldots, \hat{G}(n)\right)$. Here we think of $\hat{G}(n)$ as a random variable with $(\lfloor \alpha n \rfloor + 2)^2$ components of the form $X_{ij}(n)$, where $X_{ij}(n)$ denotes the number of edges between vertices $i$ and $j$ at time $n$.

By definition, $(M_n)_{n=0}^t$ is a martingale. The bulk of the proof resides in the following lemma, in which we bound the martingale difference $|M_n - M_{n-1}|$. Finally, we apply the Azuma-Hoeffding inequality to prove the result.

**Lemma 8.** *The martingale $(M_n)_{n=0}^t$ satisfies $|M_n - M_{n-1}| \leq 2$ almost surely and for every $n = 1 \ldots, t$.*

*Proof.* Throughout the proof we will use the two-dimensional vector $\hat{D}_t^v = (\hat{D}_{t,1}^v, \hat{D}_{t,2}^v)$ to denote the number of in-edges from community 1 and community 2, respectively, to node $v$. Note that

$$M_n = E[\hat{N}_t(i,j)|\mathcal{F}_n] = E\left[\sum_{v \in C_1(t)} \mathbb{1}(\hat{D}_t^v = (i,j))\Big| \mathcal{F}_n\right] = \sum_{v \in C_1(t)} P\left(\hat{D}_t^v = (i,j)\,|\mathcal{F}_n\right). \quad (3.11)$$

Similarly,

$$M_{n-1} = \sum_{v \in C_1(t)} P\left(\hat{D}_t^v = (i,j)\,|\mathcal{F}_{n-1}\right)$$

For $n \geq 1$, define another graph sequence, $\tilde{G}(s)$, such that $\tilde{G}(s) = \hat{G}(s)$ for $s \leq n-1$. For $s \geq n$, $\tilde{G}(s)$ evolves independently of $\hat{G}(s)$ but according to the same evolution rules. Thus the two processes have the same marginal distribution and agree up to time $n-1$, but may disagree after.

Let $\tilde{D}_t^v$ be the in-degree of node $v$ in $\tilde{G}(t)$. Then

$$M_{n-1} = \sum_{v \in C_1(t)} P(\tilde{D}_t^v = (i,j)|\mathcal{F}_n), \quad (3.12)$$

because $\hat{G}(n)$ provides no more information than $\hat{G}(n-1) = \tilde{G}(n-1)$. Taking the difference between (3.11) and (3.12), we have

$$M_n - M_{n-1} = \sum_{v \in C_1(t)} \left[ P(\hat{D}_t^v = (i,j) | \mathcal{F}_n) - P(\tilde{D}_t^v = (i,j) | \mathcal{F}_n) \right], \tag{3.13}$$

where, conveniently, we now have two terms conditioned on the same graph. Next note that the future degree $\hat{D}_t^v$ depends only on its previous degree $\hat{D}_n^v$, and not any other features of the graph $\hat{G}(n)$, since its probability of increasing (in either coordinate) at any time $s$, $n < s \leq t$ is equal to $q\frac{\hat{D}_{s,1}^v + \hat{D}_{s,2}^v + \delta}{\lfloor qs \rfloor + \delta \lfloor \alpha s \rfloor + 2\delta}$. Therefore we can write

$$P(\hat{D}_t^v = (i,j) | \mathcal{F}_n) = P(\hat{D}_t^v = (i,j) \mid \hat{D}_n^v) \tag{3.14}$$

and

$$P(\tilde{D}_t^v = (i,j) | \mathcal{F}_n) = E\left[ P\left( \tilde{D}_t^v = (i,j) | \tilde{D}_n^v \right) \mid \mathcal{F}_n \right]. \tag{3.15}$$

In order to make sense of the conditioning in (3.14) and (3.15), we operate under the convention that if the node $v$ does not yet exist at time $n$, then $\hat{D}_n^v = (0,0)$.

Since $\hat{D}_n^v$ is measurable with respect $\mathcal{F}_n$, we can also write

$$P(\hat{D}_t^v = (i,j) | \mathcal{F}_n) = E\left[ P\left( \hat{D}_t^v = (i,j) | \hat{D}_n^v \right) \mid \mathcal{F}_n \right]. \tag{3.16}$$

Rewriting (3.13) using (3.15) and (3.16),

$$M_n - M_{n-1} = \sum_{v \in C_1(t)} E\left[ P\left( \hat{D}_t^v = (i,j) | \hat{D}_n^v \right) - P\left( \tilde{D}_t^v = (i,j) | \tilde{D}_n^v \right) \,\middle|\, \mathcal{F}_n \right]. \tag{3.17}$$

Because the two graphs evolve according to the same rules, it is true that $P(\hat{D}_t^v = (i,j) | \hat{D}_n^v) = P(\tilde{D}_t^v = (i,j) | \tilde{D}_n^v)$ whenever $\hat{D}_n^v = \tilde{D}_n^v$. Therefore

$$\left| P\left( \hat{D}_t^v = (i,j) | \hat{D}_n^v \right) - P\left( \tilde{D}_t^v = (i,j) | \tilde{D}_n^v \right) \right| \leq \mathbb{1}(\hat{D}_n^v \neq \tilde{D}_n^v). \tag{3.18}$$

We apply (3.18) to the summand in (3.17) to find that

$$|M_n - M_{n-1}| \leq E\left[\sum_{v \in C_1(t)} \mathbb{1}(\hat{D}_n^v \neq \tilde{D}_n^v) \,|\mathcal{F}_n\right].$$

Recall that $\hat{D}_{n-1}^v = \tilde{D}_{n-1}^v$ for every $v \in C_1(n-1)$ by construction. At time $n$ we add at most 1 edge in both $\hat{G}(n)$ and $\tilde{G}(n)$, so there are at most two $v$'s where $\hat{D}_n^v \neq \tilde{D}_n^v$. Thus

$$\sum_{v \in C_1(t)} \mathbb{1}(\hat{D}_n^v \neq \tilde{D}_n^v) \leq 2,$$

and we conclude that $|M_n - M_{n-1}| \leq 2$. □

Now we are prepared to prove Lemma 6. The proof is a straight forward application of the Azuma-Hoeffding inequality to the martingale $(M_n)_{n=0}^t$.

*Proof of Lemma 6.* We know that there are exactly $\lfloor qt \rfloor \leq t$ edges in $\hat{G}(t)$. Thus we can reduce the proof by noticing that $N_t(i,j) = 0$ for all $(i,j)$ such that $i + j > t$. Hence we are looking to prove

$$P\left(\max_{i,j\,:\,i+j\leq t} \left|\hat{N}_t(i,j) - E[\hat{N}_t(i,j)]\right| \geq C\sqrt{t\log t}\right) = o(1).$$

We begin by bounding the maximum by the sum over all $i, j$.

$$\begin{aligned} &P\left(\max_{i,j\,:\,i+j\leq t} |\hat{N}_t(i,j) - E[\hat{N}_t(i,j)]| \geq C\sqrt{t\log t}\right) \\ &\quad\leq \sum_{i=0}^{t}\sum_{j=0}^{t-i} P\left(\left|\hat{N}_t(i,j) - E[\hat{N}_t(i,j)]\right| \geq C\sqrt{t\log t}\right). \end{aligned}$$

Since $|M_n - M_{n-1}| \leq 2$ for all $n$ by Lemma 8, the Azuma-Hoeffding inequality states that for any $a > 0$,

$$P\left(|M_t - EM_t| \geq a\right) \leq 2\exp\left\{-\frac{a^2}{2\sum_{i=1}^t 2^2}\right\}.$$

Note $M_t = \hat{N}_t(i,j)$. Taking $a = C\sqrt{t \log t}$ for $C > 4$, we have

$$P\left(|\hat{N}_t(i,j) - E\hat{N}_t(i,j)| \geq C\sqrt{t \log t}\right) \leq 2\exp\left\{-\frac{C^2 t \log t}{8t}\right\} = 2t^{-C^2/8} = o(t^{-2}).$$

Summing over $0 \leq i + j \leq t$ proves the lemma.

□

### 3.3.3 Expected degree distribution

We have now arrived at the final component. In Section 3.3.1, we showed that, $N_t(i,j)$, the number of nodes with degree $(i,j)$ in $G(t)$, is close to $\hat{N}_t(i,j)$, the corresponding quantity in $\hat{G}(t)$. Thus we need only find the degree distribution in $\hat{G}(t)$. The martingale argument of Section 3.3.2 shows that $\hat{N}_t(i,j)$ concentrates around its expectation, $E[\hat{N}_t(i,j)]$. Finally, we will show that $E[\hat{N}_t(i,j)]$ looks like $\frac{\alpha}{2}tp(i,j)$, which then implies that the expected proportion of nodes with degree $(i,j)$ converges to $p(i,j)$.

We will use an induction argument to prove Lemma 7. Suppose that $E[\hat{N}_t(i,j)]$ is close to $\frac{\alpha}{2}tp(i,j)$ for some $t$. Then, in the next time step, the difference cannot increase by too much. On average, we will see that the expected increase is proportional to $p(i,j)$. In the remainder of this section, we consider how the graph sequence $\hat{G}$ evolves in one time step. We set up certain notations and relationships which will be assembled in the induction proof, which can be found in Appendix B.

Let us fix the values of $(i,j)$ and consider how $\hat{N}_t(i,j)$ could change in the one time step. The difference $\hat{N}_{t+1}(i,j) - \hat{N}_t(i,j)$ will be either -1, 0, or 1. The difference will be negative if a node with in-degree $(i,j)$ gains a new edge, because its

degree will then be altered, and the total number of nodes with degree $(i, j)$ will decrease by 1. Alternatively, the difference will be positive if a node with degree $(i-1, j)$ gains a new edge from community 1 (for $i \geq 1$), or if a node with degree $(i, j-1)$ gains a new edge from community 2 (for $j \geq 1$). If $i = j = 0$, then $\hat{N}_t(0,0)$ will increaes by 1 if a new node arrives and joins community 1. If none of the previous actions occur, then there is no change and so the difference $\hat{N}_{t+1}(i, j) - \hat{N}_t(i, j)$ will be 0.

Next we will find the probabilities of each of these possible actions, but first let us set up some notation. We note that $\hat{G}(t)$ has exactly $\lfloor \alpha t/2 \rfloor + 1$ nodes in community 1, and we let $\hat{m}_1(t)$ denote the number of community 1 edges in $\hat{G}(t)$. Since arrivals of nodes and edges are deterministic in $\hat{G}$, denote

$$\mathbb{1}_{\alpha,t} = \mathbb{1}(\lfloor \alpha(t+1) \rfloor > \lfloor \alpha t \rfloor)$$

$$\mathbb{1}_{q,t} = \mathbb{1}(\lfloor q(t+1) \rfloor > \lfloor q t \rfloor)$$

$$\mathbb{1}_{\text{odd},t} = \mathbb{1}(\lfloor \alpha(t+1) \rfloor \text{ is odd}).$$

Thus, for example, $\mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t} = 1$ implies that a new community 1 node is added to $\hat{G}(t)$.

The probability that a node with degree $(i, j)$ gains a new edge at time $t+1$ is

$$\mathbb{1}_{q,t}\left(\frac{q_{with}}{2q}\mathbb{1}_{\text{odd},t} + \frac{q_{bet}}{2q}(1 - \mathbb{1}_{\text{odd},t})\right)\frac{i + j + \delta}{\hat{m}_1(t) + \delta\lfloor \alpha t/2 \rfloor + \delta}\hat{N}_t(i, j).$$

Recall that $\hat{N}_t(i, j)$ is defined to be 0 if $i < 0$ or $j < 0$. Then the probability that a node with degree $(i-1, j)$ gains an edge from community 1 at time $t+1$ is

$$\mathbb{1}_{q,t}\left(\frac{q_{with}}{2q}\mathbb{1}_{\text{odd},t}\right)\frac{i + j - 1 + \delta}{\hat{m}_1(t) + \delta\lfloor \alpha t/2 \rfloor + \delta}\hat{N}_t(i-1, j),$$

and the probability that a node with degree $(i, j-1)$ gains an edge from a community 2 is

$$\mathbb{1}_{q,t}\frac{q_{bet}}{2q}(1 - \mathbb{1}_{\text{odd},t})\frac{i + j - 1 + \delta}{\hat{m}_1(t) + \delta\lfloor \alpha t/2 \rfloor + \delta}\hat{N}_t(i, j-1).$$

Therefore, by conditioning upon all information up to and including time $t$, we can write

$$
\begin{aligned}
E[\hat{N}_{t+1}(i,j) - \hat{N}_t(i,j) \mid \mathcal{F}_t] &= \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t}\mathbb{1}(i=j=0) \\
&+ \mathbb{1}_{q,t}\frac{i+j-1+\delta}{\hat{m}_1(t)+\delta\lfloor\alpha t/2\rfloor+\delta}\left(\frac{q_{with}}{2q}\mathbb{1}_{\text{odd},t}\hat{N}_t(i-1,j) + \frac{q_{bet}}{2q}\left(1-\mathbb{1}_{\text{odd},t}\right)\hat{N}_t(i,j-1)\right) \\
&- \mathbb{1}_{q,t}\frac{i+j+\delta}{\hat{m}_1(t)+\delta\lfloor\alpha t/2\rfloor+\delta}\left(\frac{q_{with}}{2q}\mathbb{1}_{\text{odd},t} + \frac{q_{bet}}{2q}\left(1-\mathbb{1}_{\text{odd},t}\right)\right)\hat{N}_t(i,j).
\end{aligned} \tag{3.19}
$$

Taking another expectation, we remove the conditional and transform (3.19) into the more useful form

$$
\begin{aligned}
E[\hat{N}_{t+1}(i,j)] &= \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t}\mathbb{1}(i=j=0) + \mathbb{1}_{q,t}E[\Delta_t(i,j)] \\
&+ \left(1 - \mathbb{1}_{q,t}\frac{i+j+\delta}{\frac{1}{2}(q+\alpha\delta)t}\left(\frac{q_{with}}{2q}\mathbb{1}_{\text{odd},t} + \frac{q_{bet}}{2q}\left(1-\mathbb{1}_{\text{odd},t}\right)\right)\right)E[\hat{N}_t(i,j)] \\
&+ \mathbb{1}_{q,t}\frac{i+j-1+\delta}{\frac{1}{2}(q+\alpha\delta)t}\left(\frac{q_{with}}{2q}\mathbb{1}_{\text{odd},t}E[\hat{N}_t(i-1,j)] + \frac{q_{bet}}{2q}(1-\mathbb{1}_{\text{odd},t})E[\hat{N}_t(i,j-1)]\right),
\end{aligned} \tag{3.20}
$$

where

$$
\begin{aligned}
\Delta_t(i,j) = &\left[\frac{1}{\hat{m}_1(t)+\delta\lfloor\alpha t/2\rfloor+\delta} - \frac{1}{\frac{1}{2}(q+\alpha\delta)t}\right] \\
&\cdot\left[(i+j-1+\delta)\left(\frac{q_{with}}{2q}\mathbb{1}_{\text{odd},t}\hat{N}_t(i-1,j) + \frac{q_{bet}}{2q}(1-\mathbb{1}_{\text{odd},t})\hat{N}_t(i,j-1)\right)\right. \\
&\left.-(i+j+\delta)\left(\frac{q_{with}}{2q}\mathbb{1}_{\text{odd},t} + \frac{q_{bet}}{2q}\left(1-\mathbb{1}_{\text{odd},t}\right)\right)\hat{N}_t(i,j)\right].
\end{aligned} \tag{3.21}
$$

We will show in Lemma 9 (below), that $E|\Delta_t(i,j)|$ decays like $1/t$.

**Lemma 9.** *For all $t \geq 1$ and any $i, j \geq 0$,*

$$
E|\Delta_t(i,j)| \leq \frac{T+\delta}{\frac{1}{2}(q+\alpha\delta)}t^{-1}.
$$

*Proof.* By (3.21),

$$
\begin{aligned}
|\Delta_t(i,j)| &\leq \left|\frac{1}{\frac{1}{2}(q+\alpha\delta)t} - \frac{1}{\hat{m}_1(t)+\delta\lfloor\alpha t/2\rfloor+\delta}\right|\sum_{i,j}(i+j+\delta)\hat{N}_t(i,j) \\
&= \left|\frac{(\hat{m}_1(t)-\frac{q}{2}t)+\delta(\lfloor\alpha t/2\rfloor+1-\alpha t/2)}{\left(\frac{1}{2}(q+\alpha\delta)t\right)(\hat{m}_1(t)+\delta\lfloor\alpha t/2\rfloor+\delta)}\right|\sum_{i,j}(i+j+\delta)\hat{N}_t(i,j).
\end{aligned}
$$

Since $\sum_{i,j}(i+j+\delta)\hat{N}_t(i,j) = \hat{m}_1(t) + \delta\lfloor \alpha t/2 \rfloor + \delta$, we have

$$E|\Delta_t(i,j)| \leq E\left|\frac{(\hat{m}_1(t) - \frac{q}{2}t) + \delta(\lfloor \alpha t/2 \rfloor + 1 - \alpha t/2)}{\frac{1}{2}(q+\alpha\delta)t}\right|,$$

and by Lemma 23 we know that the number of community 1 edges is always close to $\frac{1}{2}qt$, so

$$E|\Delta_t(i,j)| \leq \frac{E|\hat{m}_1(t) - \frac{q}{2}t| + \delta}{\frac{1}{2}(q+\alpha\delta)t} \leq \frac{T+\delta}{\frac{1}{2}(q+\alpha\delta)}t^{-1}.$$

□

The gist of the proof of Lemma 7 lies in comparing the expansion of $E[\hat{N}_{t+1}(i,j)]$, as in (3.20), to the $p(i,j)$ recursion (3.4). We will see that these recursions are quite similar, and therefore that the proportion of nodes with degree $(i,j)$ is approximately equal to $p(i,j)$. The difficulty in the comparison is the presence of the indicator functions in (3.20). The indicator functions $\mathbb{1}_{q,t}$, $\mathbb{1}_{\alpha,t}$, and $\mathbb{1}_{\text{odd},t}$ occur periodically, about every $\frac{1}{q}$, $\frac{1}{\alpha}$, and 2 timesteps, respectively. Thus the proof relies on the period $T$ defined in Section 3.3. The details of the proof are given in Appendix B.

This concludes the last step in the proof of our main result, Theorem 14, stating that the proportion of community 1 nodes with in-degree $(i,j)$ in $G(t)$ converges to $p(i,j)$ as $t \to \infty$. In the following section, we will restate and prove Theorem 13 in the special case of two symmetric communities. This theorem gives the regular variation result of the $p(i,j)$ distribution.

## 3.4 Regular Variation in the $k = 2$, symmetric case

In this section, we prove Theorem 13 in the case of two symmetric communities described in Section 3.3. This will show that the distribution $p(i,j)$ given in (3.5)

is regularly varying in the multivariate sense. We will see that the tail index is $1 + \frac{\alpha\delta}{q}$, where $q = \frac{1}{2}(q_{with} + q_{bet})$. We will also see that the spectral measure of this distribution concentrates around the line defined by the parametric equation $(q_{with}, q_{bet})t, t > 0$.

**Theorem 15.** *Let $(W, B)$ be a random vector from distribution $(p(i, j) : i, j \geq 0)$. Let $c = \frac{\alpha\delta}{q}$ and let $\nu(\cdot)$ be the measure concentrating on the ray $\{(q_{with}, q_{bet}t : t > 0\}$, such that the set*

$$A = \left\{ X \in \mathbb{R}^2 : \|X\| > a, \frac{X}{\|X\|} = \frac{(q_{with}, q_{bet})}{\sqrt{q_{with}^2 + q_{bet}^2}} \right\}$$

*has measure*

$$\nu(A) = a^{-(c+1)} \frac{\Gamma(1 + c + \delta)}{\Gamma(\delta)} \left( \frac{\sqrt{q_{with}^2 + q_{bet}^2}}{2q} \right)^{c+1}.$$

*As $x \to \infty$,*

$$x^{c+1} P\left( \frac{1}{x}(W, B) \in \cdot \right) \xrightarrow{v} \nu(\cdot).$$

*Proof.* Fix $0 < \theta_1 < \theta_2$ and let $K$ be defined by $K = \{(x, y) : x^2 + y^2 > 1, \theta_1 x \leq y \leq \theta_2 x\}$. Then

$$P\left( \frac{1}{x}(W, B) \in K \right) = P\left( W > \frac{x}{\sqrt{1 + \theta_1^2}}, \theta_1 W < B < \theta_2 W \right) = \sum_{i = x/\sqrt{1+\theta_1^2}}^{\infty} \sum_{j=\theta_1 i}^{\theta_2 i} p(i, j).$$

As $i, j \to \infty$, the sum is approximated by an appropriate integral. Additionally, applying Stirling's formula to (3.5) for large $i, j$, we can write

$$P\left( \|(W, B)\| > x, \frac{B}{W} \in (\theta_1, \theta_2) \right) \sim \frac{(1 + c)\Gamma(1 + c + \delta)}{\Gamma(\delta)}$$
$$\cdot \int_{x/\sqrt{1+\theta_1^2}}^{\infty} \int_{\theta_1 w}^{\theta_2 w} \frac{1}{\sqrt{2\pi}} \frac{\sqrt{w + b}}{\sqrt{w}\sqrt{b}} \frac{(w + b)^{w+b}}{w^w b^b} \left( \frac{q_{with}}{2q} \right)^w \left( \frac{q_{bet}}{2q} \right)^b (w + b)^{-(c+2)} \, dbdw$$

as $x \to \infty$. With a change of variables $b = wy$, the inner integral becomes

$$\frac{1}{\sqrt{2\pi}} w^{-(c+\frac{3}{2})} \int_{\theta_1}^{\theta_2} y^{-\frac{1}{2}} (1 + y)^{-(c+\frac{3}{2})} \left[ \left( \frac{q_{with}}{2q} \right) \left( \frac{q_{bet}}{2q} \right)^y \frac{(1 + y)^{1+y}}{y^y} \right]^w dy. \tag{3.22}$$

Define

$$g(\theta_1, \theta_2) = \max_{y \in (\theta_1, \theta_2)} \left[ \left( \frac{q_{with}}{2q} \right) \left( \frac{q_{bet}}{2q} \right)^y \frac{(1+y)^{1+y}}{y^y} \right]$$

and note that $g(\theta_1, \theta_2) \leq 1$, with equality only when $\frac{q_{bet}}{q_{with}} \in (\theta_1, \theta_2)$. If $\frac{q_{bet}}{q_{with}} \notin (\theta_1, \theta_2)$, it follows that

$$\begin{aligned} & x^{c+1} P\left( \|(W, B)\| > x, \frac{B}{W} \in (\theta_1, \theta_2) \right) \\ & \leq x^{c+1} \frac{(1+c)\Gamma(1+c+\delta)}{\sqrt{2\pi}\Gamma(\delta)} \int_{x/\sqrt{1+\theta_1^2}}^{\infty} \theta_1^{-\frac{1}{2}} (1+\theta_1)^{-(c+\frac{3}{2})} g(\theta_1, \theta_2)^w (\theta_2 - \theta_1)\, dw \\ & \leq K x^{c+1} \left[ g(\theta_1, \theta_2) \right]^{x/\sqrt{1+\theta_1^2}} \\ & \to 0 \end{aligned} \tag{3.23}$$

where

$$K = -\frac{(1+c)\Gamma(1+c+\delta)}{\sqrt{2\pi}\Gamma(\delta)} \frac{\theta_2 - \theta_1}{\sqrt{\theta_1}(1+\theta_1)^{c+\frac{3}{2}}} \left( \frac{1}{\log g(\theta_1, \theta_2)} \right).$$

For the case $\frac{q_{bet}}{q_{with}} \in (\theta_1, \theta_2)$, denote $\theta = \frac{q_{bet}}{q_{with}}$, and consider, with fixed $i$, (3.22) over a small interval containing $\theta$.

$$\sum_{j=(1-\epsilon)\theta i}^{(1+\epsilon)\theta i} p(i,j) \sim i^{-(c+\frac{3}{2})} \frac{1}{\sqrt{2\pi}} \int_{(1-\epsilon)\theta}^{(1+\epsilon)\theta} y^{-\frac{1}{2}} (1+y)^{-(c+\frac{3}{2})} \left[ \frac{(1+y)^{1+y}}{y^y} \frac{\theta^y}{(\theta+1)^{1+y}} \right]^i dy,$$

Where we've used that

$$\frac{1}{\theta+1} = \frac{q_{with}}{2q}, \qquad \frac{\theta}{\theta+1} = \frac{q_{bet}}{2q}.$$

Define the function

$$h(y) = (1+y)\log(1+y) - y\log y + y\log\theta - (1+y)\log(\theta+1)$$

so that

$$\sum_{j=(1-\epsilon)\theta i}^{(1+\epsilon)\theta i} p(i,j) \sim i^{-(c+\frac{3}{2})} \frac{(\theta+1)^{-(c+\frac{3}{2})}}{\sqrt{2\pi\theta}} \int_{(1-\epsilon)\theta}^{(1+\epsilon)\theta} e^{ih(y)}\, dy.$$

Note $h$ is maximized at $y = \theta$, with $h(\theta) = 0$. A second-order Taylor approximation about this point and a change of variables $z = (y - \theta)\sqrt{i/2\theta(\theta+1)}$ shows

$$\sum_{j=(1-\epsilon)\theta i}^{(1+\epsilon)\theta i} p(i,j) \sim i^{-(c+2)} \frac{2}{\sqrt{\pi}} (\theta+1)^{-(c+1)} \int_0^\infty e^{-z^2}\, dz$$

as $i \to \infty$. Then the fact $\int_0^\infty e^{-z^2} ds = \sqrt{\pi}/2$ yields

$$\sum_{j=(1-\epsilon)\theta i}^{(1+\epsilon)\theta i} p(i,j) \sim (\theta+1)^{-(c+1)} i^{-(c+2)}. \tag{3.24}$$

Finally, from (3.23) and (3.24),

$$\begin{aligned}
&\lim_{x\to\infty} x^{c+1} P\left(\|(W,B)\| > x, \frac{B}{W} \in (\theta_1, \theta_2)\right) \\
&\quad = \lim_{x\to\infty} x^{c+1} P\left(\|(W,B)\| > x, \frac{B}{W} \in ((1-\epsilon)\theta, (1+\epsilon)\theta)\right) \\
&\quad \lim_{x\to\infty} x^{c+1} \frac{(1+c)\Gamma(1+c+\delta)}{\Gamma(\delta)} \int_{x/\sqrt{1+\theta^2}}^\infty (\theta+1)^{-(c+1)} t^{-(c+2)}\, dt \\
&\quad = \frac{\Gamma(1+c+\delta)}{\Gamma(\delta)} \left(\frac{\sqrt{q_{bet}^2 + q_{with}^2}}{2q}\right)^{c+1}.
\end{aligned}$$

□

Inuitively, for a node with total in-degree $d$, one might expect the proportion of within-edges to be about $\frac{q_{with}}{2q}$ and the proportion of between-edges to be about $\frac{q_{bet}}{2q}$. Theorem 15 states that these ratios are observed in the tails, that is, for $d$ very large. This is illustrated in Figure 3.1, where within-degree is plotted against between-degree. From Theorem 15, we expect the points to fall near the line through the origin with slope $q_{with}/q_{bet}$. When only observing the lower degree nodes (left), we see a trend, but not clear adherence to this line. However, as we include the few higher degree nodes (right), the expected convergence emerges.

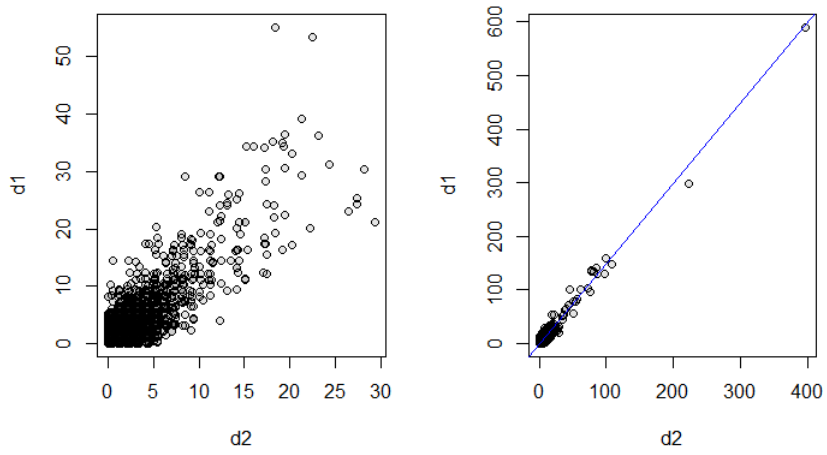



Figure 3.1: From a simulation of 100,000 time steps with $\alpha = 0.5$, $\delta = 0.5$, $q_{with} = 0.6$, $q_{bet} = 0.4$, there are 24816 Community 1 nodes. On the left, we plot in-degree coming from Community 1 ($d_1$) vs. in-degree from Community 2 ($d_2$) for all except the 26 highest degree nodes. On the right, we plot all nodes and compare to the line $(q_{with}, q_{bet})t$, $t \geq 0$ (in blue).

## 3.5 General $k$

We now consider the general case, where the number of communities $k$ may be larger than 2, and the matrix of acceptance probabilities $q_{ij}$, $i, j = 1, \ldots, k$ may not be symmetric. In Section 3.5.1, we prove Theorem 13 for the $k$ dimensional case. This result gives the regular variation result for $p_i(\cdot)$, as defined in (3.1).

In Section 3.5.2, we give the intuition behind the proof of Theorem 12, which states that the fraction $P_{i,t}(\mathbf{d})$ of community 1 nodes with degree $\mathbf{d}$ at time $t$ converges to the $p_i(\mathbf{d})$. The convergence proof for general $k$ requires cumbersome notation but does not provide additional insight over the $k = 2$ case, and is omitted. However, we do include a short sketch to suggest how this convergence comes about, and to explain where the recursion comes from.

### 3.5.1 Regular Variation for General $k$

In this section, we prove Theorem 13 for the general case, which is restated below.

Consider the random vector $\mathbf{D_i} = (D_1, \ldots, D_k)$, which has distribution

$$P(\mathbf{D_i} = \mathbf{d}) = \frac{k}{\alpha} p_i(\mathbf{d}),$$

for $i \in \{1, \ldots, k\}$ fixed.

**Theorem 16.** *There exists a radon measure $\nu_i(\cdot)$ concentrating on the ray $((q_{1i}, q_{2i}, \ldots, q_{ki})t, t > 0)$ in $\mathbb{R}^k$ such that*

$$x^{c+1} P\left(\frac{1}{x}\mathbf{D} \in \cdot\right) \xrightarrow{v} \nu^i(\cdot).$$

*Specifically, the set*

$$A = \left\{ \mathbf{d} \in \mathbb{R}^k : \|\mathbf{d}\| > a, \frac{\mathbf{d}}{\|\mathbf{d}\|} = \frac{(q_{1i}, \ldots, q_{ki})}{\sqrt{q_{1i}^2 + \cdots + q_{ki}^2}} \right\}$$

*has measure $\nu_i(A) = c_i a^{-(c+1)}$.*

*Proof.* Let $K$ be a cone in the positive orthant of $\mathbb{R}^k$ but with the unit cube removed, such that we can write

$$P\left(\frac{1}{x}\mathbf{D_i} \in K\right) = P\left(D_1 > \frac{x}{r},\ a_j D_1 < D_j < b_j D_1,\ j = 2, \ldots, k\right)$$

for some $0 < a_j < b_j$, $j = 2, \ldots, k$, $r = \sqrt{1 + a_2^2 + \cdots + a_k^2}$.

Then as $x \to \infty$,

$$\sum_{(d_1, d_2, \ldots, d_k) \in K} p^i(\mathbf{d}) \sim \int_{x/r}^{\infty} \int_{a_2}^{b_2} \cdots \int_{a_k}^{b_k} p^i\left((1, y_2, y_3, \ldots, y_k)x_1\right) x_1^{k-1}\, dy_k \cdots dy_2 dx_1.$$

From Stirling's formula, we expand this as

$$\frac{c+1}{(2\pi)^{(k-1)/2}}\frac{\Gamma(1+c+\delta)}{\Gamma(\delta)}\int_{x/r}^{\infty} x_1^{-c-\frac{5}{2}+\frac{k}{2}}\int_{a_2}^{b_2}\cdots\int_{a_k}^{b_k}\frac{(1+y_2+\cdots+y_k)^{-(c+\frac{3}{2})}}{\sqrt{y_2\cdots y_k}}$$
$$\cdot\left[\left(\frac{1+y_2+\cdots+y_k}{kq}\right)^{1+y_2+\cdots y_k} q_{1i}\left(\frac{q_{2i}}{y_2}\right)^{y_2}\cdots\left(\frac{q_{ki}}{y_k}\right)^{y_k}\right]^{x_1} dy_k\cdots dy_2 dx_1.$$

Define the function

$$g(K)=\max_{y_j\in(a_j,b_j),j=2,\ldots,k}\left[\left(\frac{1+y_2+\cdots+y_k}{kq}\right)^{1+y_2+\cdots y_k} q_{1i}\left(\frac{q_{2i}}{y_2}\right)^{y_2}\cdots\left(\frac{q_{ki}}{y_k}\right)^{y_k}\right].$$

Notice $g(K)\le 1$, with equality only if $K$ contains the line $\frac{1}{q_{1i}}(q_{1i},\ldots,q_{ki})x_1$.

In the case that $K$ does not contain this line, then

$$P\left(\frac{1}{x}\mathbf{D_i}\in K\right)\le -(c+1)\left(\frac{2}{\pi}\right)^{\frac{k-1}{2}}\frac{\Gamma(1+c+\delta)}{\Gamma(\delta)}\frac{(\sqrt{b_k}-\sqrt{a_k})\cdots(\sqrt{b_2}-\sqrt{a_2})}{\log(g(K))}g(K)^{x/r}.$$

Thus since $g(K)<1$ in this case,

$$\lim_{x\to\infty} x^{c+1}P\left(\frac{1}{x}\mathbf{D_i}\in K\right)=\lim_{x\to\infty} C_K g(K)^{x/r}x^{c+1}=0.$$

Now let's return to the case where $K$ does contain the line $\frac{1}{q_{1i}}(q_{1i},\ldots,q_{ki})x_1$. Then the tail probability is asymptotically equal to the integral over a small region surrounding the line. Specifically, we take the bounds of integration to be $a_j=(1-\epsilon)\frac{q_{ji}}{q_{1i}}$, $b_j=(1+\epsilon)\frac{q_{ji}}{q_{1i}}$ for some small $\epsilon>0$, $j=2,\ldots,k$.

Define

$$h(y_2,\ldots,y_k)=(1+y_2+\cdots+y_k)\log(1+y_2+\cdots+y_k)-(1+y_2+\cdots+y_k)\log(kq)$$
$$+\log q_{1i}+y_2\log q_{2i}+\cdots+y_k\log q_{ki}-y_2\log y_2-\cdots-y_k\log y_k$$

so that we can write

$$P\left(\frac{1}{x}\mathbf{D_i}\in K\right)\sim\frac{c+1}{(2\pi)^{(k-1)/2}}\frac{\Gamma(1+c+\delta)}{\Gamma(\delta)}\int_{x/r}^{\infty} x_1^{-c-\frac{5}{2}+\frac{k}{2}}\int_{(1-\epsilon)\frac{q_{2i}}{q_{1i}}}^{(1+\epsilon)\frac{q_{2i}}{q_{1i}}}\cdots\int_{(1-\epsilon)\frac{q_{ki}}{q_{1i}}}^{(1+\epsilon)\frac{q_{ki}}{q_{1i}}}$$
$$\cdot\frac{(1+y_2+\cdots+y_k)^{-(c+\frac{3}{2})}}{\sqrt{y_2\cdots y_k}}e^{x_1 h(y_2,\ldots,y_k)}\,dy_k\cdots dy_2 dx_1.$$

Notice that $h(\frac{1}{q_{1i}}(q_{2i},\dots,q_{ki})) = 0$ and $h$ is maximized at this point. Using a second-order Taylor approximation about this point, and with the appropriate change of variables, we can write

$$P\left(\frac{1}{x}\mathbf{D_i}\in K\right) \sim \frac{c+1}{\pi^{\frac{k-1}{2}}}\frac{\Gamma(1+c+\delta)}{\Gamma(\delta)}\frac{2^{k-1}J}{(kq)^{c+2-\frac{k}{2}}}\frac{q_{1i}^{c+\frac{3}{2}}}{(q_{2i}\cdots q_{ki})^{\frac{1}{2}}}\int_{x/r}^{\infty} x_1^{-(c+2)} \cdot \int_0^{\epsilon a_2\sqrt{\frac{x_1}{2q_{1i}kq}}}\cdots\int_0^{\epsilon a_k\sqrt{\frac{x_1}{2q_{1i}kq}}} e^{\sum_{j=2}^k \lambda_j z_j^2} dz_k\cdots dz_2 dx_1,$$

where $\lambda_2,\dots,\lambda_k<0$ are the eigenvalues of the Hessian matrix for $h$, and $J$ is the determinant of the $(k-1)\times(k-1)$ matrix $\left[(\frac{kq}{q_{\ell i}}+\lambda_j)^{-1}\right]_{j,\ell=2}^{k}$. As $x\to\infty$, we use that $\int_0^\infty e^{-z^2}dz = \frac{\sqrt{\pi}}{2}$ to get

$$P\left(\frac{1}{x}\mathbf{D_i}\in K\right) \sim \frac{c+1}{\pi^{\frac{k-1}{2}}}\frac{\Gamma(1+c+\delta)}{\Gamma(\delta)}\frac{2^{k-1}}{(kq)^{c+2-\frac{k}{2}}}\frac{q_{1i}^{c+\frac{3}{2}}}{(q_{2i}\cdots q_{ki})^{\frac{1}{2}}}\frac{J}{(\lambda_2\cdots\lambda_k)^{\frac{1}{2}}}\int_{x/r}^{\infty} x_1^{-(c+2)}\left(\frac{\sqrt{\pi}}{2}\right)^{k-1} dx_1.$$

Finally, integrating the last term and recalling that $r=\frac{1}{q_{1i}}\sqrt{q_{1i}^2+\cdots+q_{ki}^2}$,

$$\begin{aligned}&\lim_{x\to\infty} x^{c+1}P\left(\frac{1}{x}\mathbf{D_i}\in K\right)\\ &= \frac{\Gamma(1+c+\delta)}{\Gamma(\delta)}\sqrt{\frac{q_{1i}}{kq}}\frac{J(kq)^{\frac{k-1}{2}}}{\sqrt{(\lambda_2 q_{2i})\cdots(\lambda_k q_{ki})}}\left(\frac{\sqrt{q_{1i}^2+\cdots+q_{ki}^2}}{kq}\right)^{c+1}.\end{aligned}$$

This proves Theorem 16, with the constant $c_i$ given by

$$c_i = \frac{\Gamma(1+c+\delta)}{\Gamma(\delta)}\sqrt{\frac{q_{1i}}{kq}}\frac{J(kq)^{\frac{k-1}{2}}}{\sqrt{(\lambda_2 q_{2i})\cdots(\lambda_k q_{ki})}}\left(\frac{\sqrt{q_{1i}^2+\cdots+q_{ki}^2}}{kq}\right)^{c+1}, \tag{3.25}$$

where

$$J = \det\begin{pmatrix} \frac{1}{\frac{kq}{q_{2i}}+\lambda_2} & \frac{1}{\frac{kq}{q_{2i}}+\lambda_3} & \cdots & \frac{1}{\frac{kq}{q_{2i}}+\lambda_k}\\ \frac{1}{\frac{kq}{q_{3i}}+\lambda_2} & \frac{1}{\frac{kq}{q_{3i}}+\lambda_3} & & \\ \vdots & & \ddots & \vdots \\ \frac{1}{\frac{kq}{q_{ki}}+\lambda_2} & & \cdots & \frac{1}{\frac{kq}{q_{ki}}+\lambda_k}\end{pmatrix}.$$

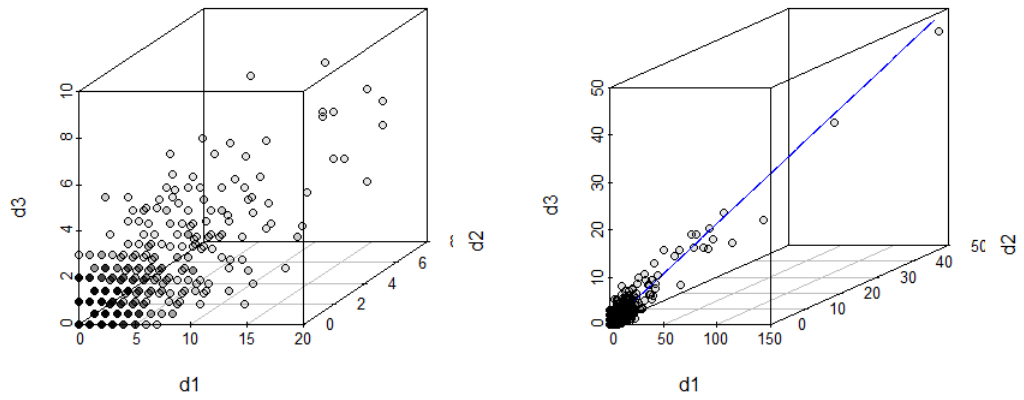



Figure 3.2: From a simulation of 50,000 timesteps with $\alpha = 0.4$, $\delta = 1$, $q_{11} = 0.9$, $q_{21} = 0.3$, $q_{31} = 0.3$, there are 6620 Community 1 nodes. Plot degrees $(d_1, d_2, d_3)$ (right), and compare to the line $(q_{11}, q_{21}, q_{31})t$, $t \geq 0$ (in blue). Removing the 20 highest degree nodes (left) shows a trend, but not a precise relationship

□

See Figure 3.2 for an illustration of this result for $k = 3$. We expect the degree vector $(d_1, d_2, d_3)$ of a community 1 node to fall along the parametric line $(q_{11}, q_{21}, q_{31})t$, $t \geq 0$ when total degree is large. In Figure 3.2 (right), we see a good match in the high degree regime.

### 3.5.2 Intuition for convergence, general $k$

Here we give the intuition behind the convergence of $P_{t,i}(\mathbf{d})$ to $p_i(\mathbf{d})$. This section is most similar to Section 3.3.3, because we consider how the graph's degree distribution can change in one time step.

Let $\mathbf{e}_j$ be the vector with a 1 in the $j$th coordinate, and 0's elsewhere. We let

$C_i(t)$ denote the set of all nodes in $G(t)$ which belong to community $i$. Considering how $N_t^i(\mathbf{d})$ may change in the next time step, we can write

$$\begin{aligned} E\left[N_{t+1}^i(\mathbf{d}) - N_t^i(\mathbf{d}) \mid \mathcal{F}\right] &= \mathbb{1}_{\mathbf{d}=\mathbf{0}} P(\text{new node in } C_i) \\ &\quad + \sum_{j=1}^{k} P(v \to w,\ v \in C_j(t),\ w \in C_i(t),\ \mathbf{D}_t^w = \mathbf{d} - \mathbf{e}_j) \\ &\quad - \sum_{j=1}^{k} P(v \to w,\ v \in C_j(t),\ w \in C_i(t),\ \mathbf{D}_t^w = \mathbf{d}). \end{aligned} \tag{3.26}$$

From the definition of our model, we calculate

$$\begin{aligned} &P(v \to w,\ v \in C_j(t),\ w \in C_i(t),\ \mathbf{D}_t^w = \mathbf{d}) \\ &\qquad = \left(\frac{\alpha}{k} + (1-\alpha)\frac{|C_j(t)|}{n(t)}\right)\left(\frac{(d_1 + d_2 + \cdots + d_k + \delta)N_t^i(\mathbf{d})}{m(t) + \delta n(t)}\right) q_{ji}. \end{aligned}$$

After plugging in these probabilities to (3.26) and some algebra, we have

$$\begin{aligned} E\left[N_{t+1}^i(\mathbf{d}) - N_t^i(\mathbf{d}) \mid \mathcal{F}\right] &= \frac{\alpha}{k}\mathbb{1}_{\mathbf{d}=\mathbf{0}} \\ &\quad + \sum_{j=1}^{k} q_{ji}\left(\frac{\alpha}{k} + (1-\alpha)\frac{|C_j(t)|}{n(t)}\right)\left(\frac{(\sum_{m=1}^k d_m - 1 + \delta)N_t^i(\mathbf{d} - \mathbf{e}_j) - (\sum_{m=1}^k d_m + \delta)N_t^i(\mathbf{d})}{m(t) + \delta n(t)}\right). \end{aligned} \tag{3.27}$$

At this point, we could include similar arguments as in Section 3.3 to prove Theorem 12 for general $k$. However, we feel that providing these arguments would not give significant additional instruction to the reader, and instead give a short heuristic argument to solidify the intuition behind these arguments.

If we assume that the proportion of nodes of degree $\mathbf{d}$ in each community converges to some proportion, we can write $\frac{N_t^i(\mathbf{d})}{n^i(t)} \to p_i(\mathbf{d})$, for some $p_i(\cdot)$, $i = 1, \ldots, k$. Additionally assuming $m(t) \approx qt$ and $n(t) \approx \alpha t$, then taking limits on both sides of (3.27) results in

$$p_i(\mathbf{d}) = \mathbb{1}_{\mathbf{d}=\mathbf{0}} + \frac{1}{k(q + \alpha\delta)} \sum_{j=1}^{k} q_{ji}\left(\sum_{m=1}^{k}(d_m - 1 + \delta)p^i(\mathbf{d} - \mathbf{e}_j) - \sum_{m=1}^{k}(d_m + \delta)p^i(\mathbf{d})\right).$$

Which, after some algebra, can be seen to be equivalent to (3.1) and hence solve (3.2).

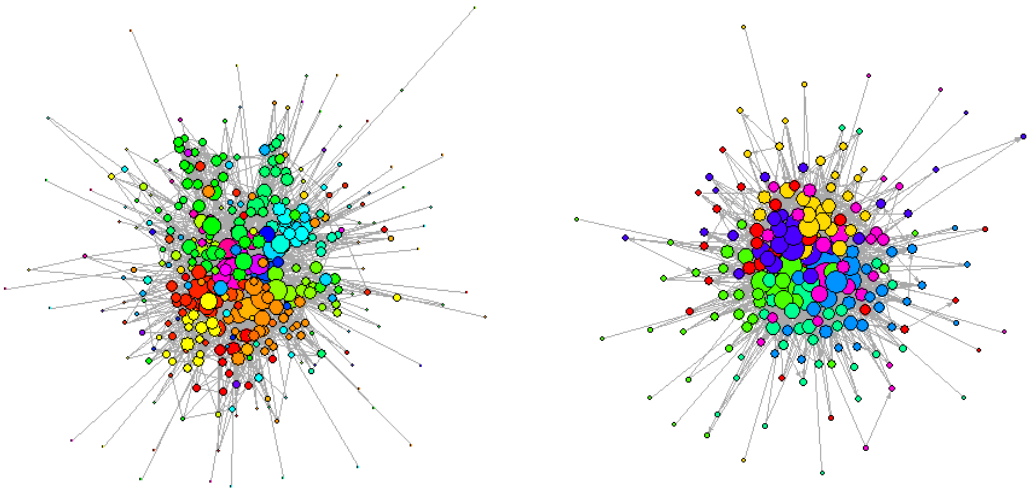
## 3.6 Conclusion

We have given a novel generative graph model with invitation and response steps, where edge invitations are sent according to the usual preferential attachment rules, and responses are made with probability depending on community membership of the nodes in question. The matrix of community-based acceptance probabilities allows the user to specify community structure, and we have shown that the asymptotic in-degree distribution still decays as a power law under this formulation. Furthermore, we have considered the joint distribution of the in-degree vector for nodes in each community and shown that it is multivariately regularly varying with tail index $1 + \frac{\alpha\delta}{q}$ corresponding to the decay of total degree. The spectral measure of the asymptotic distribution concentrates around the line given by the acceptance probabilities from each community.

The CAPAM may be a desirable model for certain applications, as it allows for a specified community structure while retaining the power law degree distribution. For example, considering the email network (shown again in Figure 3.3a), we may be able to estimate the CAPAM parameters to produce a similar-looking graph (see Figure 3.3b).

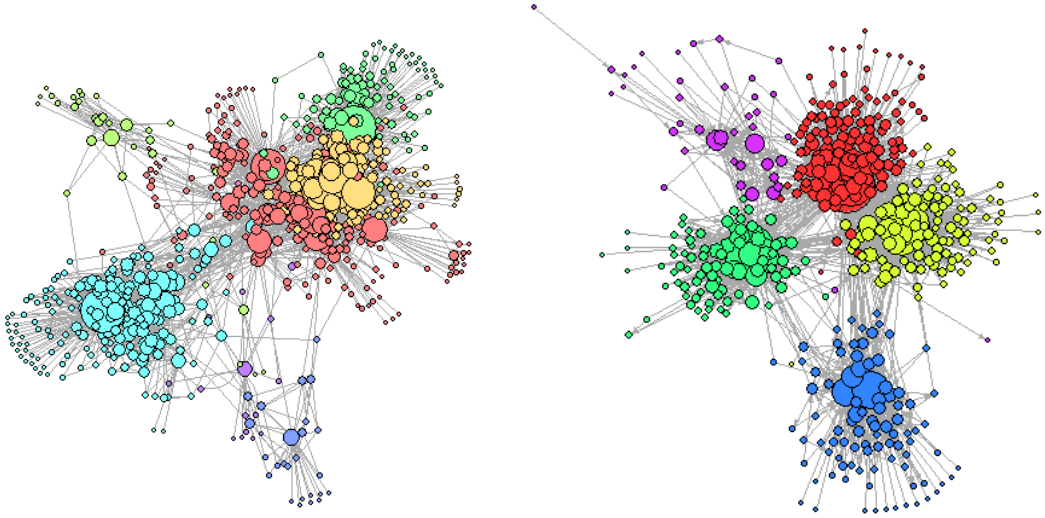
Although our formulation creates communities of equal size, we can adapt the model to create communities of various sizes. Adjusting the acceptance probability matrix can lead to more or less separation between communities (see Figure 3.4).

Our proof methods rely on a unique coupling between the CAPAM and a model in which it is known a priori the number of out-edges from each commu-



(a) (b)

Figure 3.3: (a) Email network by [92, 56] (b) CAPAM graph with $\alpha = 0.01$, $k = 7$ communities of equal size, $q_{ii} = 0.9$, $q_{ij} = 0.1$ for $i \neq j$, $\delta = 1$, and run for 20,000 time steps.



(a) (b)

Figure 3.4: (a) Social network on 700 nodes (b) CAPAM graph with $\alpha = 0.01$, $k = 5$ communities, $q_{ii} = 0.95$, $q_{ij} = 0.01$ for $i \neq j$, $\delta = 1$, and run for 50,000 time steps.

nity and the proportion of in-edges to each community. This result shows that although the CAPAM has more uncertainty, in the long run, degree distribution in the CAPAM model "balances out" to what we expect.

It may be interesting to consider the relationship between an invitation-response type PAM with the PAM defined by a fitness function. For example in comparing the CAPAM to a PAM with community-based multiplicative fitness as in [46], both models result in a similar power law index, although there is different dependency structure.

The importance of this work is that, in real-world invitation-response networks, edge formation is dependent on the terminal node as well as the initial node. In this work we have hypothesized that the user who sends an edge invitation has different motivation and preference than the user who eventually responds to the invitation. For instance, the sender may wish to grow their network and connect to influential nodes, whereas the responder may be more discerning with whom they connect to. There is a question of how dependent or related these two processes are. Furthermore, one may wonder if is there a benefit to separating these two process, or if it is equivalent to simply model them with a single fitness function. It is our hope that the CAPAM has shed a spark of insight, but future work on the invitation-response PAM in conjunction with studies of real-world networks is needed to properly address these questions.

CHAPTER 4

# COMMON NEIGHBORS MODEL

In both the Stochastic Block Model and the Community-Aware Preferential Attachment Model of the previous chapter, graphs are produced with ground-truth communities, and the relative strengths of communities are specified through a matrix parameter. The question arises, however, of how communities form? Others have suggested models with triangle-closing steps [41, 90], representing the fact that if two individuals share a common connection, then it is likely that the two individuals are connected as well. In fact, both Facebook and LinkedIn allow users to view the number of mutual friends, or common connections, between themselves and another user.

The Common Neighbors Model (CNM) takes this idea further, with the intuition that the more common neighbors between two individuals, the more likely they are to connect in future. This process has the tendency to form dense clusters, which is a characteristic seen in many real-world networks. Perhaps more importantly, these clusters form organically, without reliance on ground-truth communities.

In Section 4.1, we describe a generative random graph model where edges are formed between two users with probability proportional to the number of common neighbors. We will see that this model is difficult to analyze, due to the complicated feedback between forming a new edge and the change in number of common neighbors. To get around this difficulty, we propose a method for studying the common neighbors evolution and the graph evolution separately. We give a partial analysis of the common neighbors process in Section 4.2.

In Section 4.3 , we take a look at a CNM simulation and real network data, and consider the number of common neighbors between each pair of nodes, as well as the average number of common neighbors that one node shares with others. In Section 4.4, we compare the community structure of graphs generated by the CNM to real networks. Using the PAM as a baseline, we see that the CNM has higher modularity and cluster coefficient.

## 4.1 Model

In the CNM, we assume that nodes arrive according to a rate parameter $\alpha \in (0, 1]$. A new edge is formed at each time step, and a decision parameter $\beta \in [0, 1]$ determines whether the terminal node is selected uniformly at random, or with probability proportional to the number of common neighbors with the initial node.

We denote the number of common neighbors between nodes $v$ and $w$ at time $t \geq 0$ as

$$K_{vw}(t) = \sum_{u} A_{vu}(t)A_{wu}(t),$$

where $A(t)$ is the adjacency matrix of the graph at time $t$. We can write $K(t)$ as the matrix of all common neighbor values.

The model is as follows. Initiate graph $G(0)$ with a single node. Form $G(t+1)$ by adding to $G(t)$:

1. a new node $v$, with probability $\alpha$. Otherwise, let $v$ be a randomly chosen existing node.
2. The edge $(v, w)$, where $w$ is selected as follows:

- With probability $\beta$, select $w$ uniformly at random from the existing nodes
- With probability $1-\beta$, select $w$ with probability

$$\frac{K_{vw}(t)}{\sum_u K_{vu}(t)}.$$

If $K_{vu}(t) = 0$ for all $u$ at time $t$, then $w$ is selected at random.

Though the framework of this model is similar to preferential attachment, analysis is much more difficult, because the evolution of the matrix of common neighbors $K$ is more complicated than the evolution of node degrees. When a new edge $(v, w)$ is added to the graph, this updates the number of common neighbors for several pairs of nodes. In fact, given that the edge $(v, w)$ is added at time $t+1$, then for each $u \neq v, w$,

$$K_{vu}(t+1) = K_{vu}(t) + A_{wu}(t)$$

$$K_{wu}(t+1) = K_{wu}(t) + A_{vu}(t).$$

All common neighbor pairs must be updated after the addition of the new edge, and then used when selecting the terminal node of future edges. Since both $K$ and the adjacency matrix $A$ area changing, and increasing in dimension, over time, the feedback loop in this process becomes quite challenging to untangle.

One approach we have taken is to try to understand and model the common neighbors matrix $K$ independently of the graph evolution. If we can decouple the processes, then the model will become tractable. In Section 4.2 we propose a model for the evolution of the common neighbors matrix $K$. We give partial results on the distribution of $K(t)$ in this model. In future, the $K(t)$ distribution could then be used in the graph mdoel.

Before we discuss the $K$ evolution process, however, we take a moment to discuss similarity measures, such as Jaccard similarity, which are related to the number of common neighbors. The Jaccard similarity between two nodes $v$ and $w$ is defined as

$$J_{vw} = \frac{|N(v) \cap N(w)|}{|N(v) \cup N(w)|},$$

where $N(v)$ denotes the neighborhood of $v$, i.e., the set of nodes adjacent to $v$. Jaccard similarity is thus a measure of the number of common neighbors, scaled by the total number of neighbors. Letting $D_v(t)$ denote the degree of node $v$ at time $t$, assuming no multiple edges, then we can also write

$$J_{vw}(t) = \frac{K_{vw}(t)}{D_v(t) + D_w(t) - K_{vw}(t)}.$$

The Jaccard similarity is a useful measure, as it corrects, in a sense, for the fact that a high degree nodes also tend to have high common neighbors. One may consider a model where number of common neighbors is replaced with Jaccard similarity. This would certainly be an interesting model as well, however, it may be even more challenging to work with than simply the number of common neighbors $K$. It may not be the appropriate measure, as we will see in Sections 4.3 and 4.4, because Jaccard similarity has the deceptive property of giving high values to nodes with very few connections. Further, it may be to our advantage that number of common neighbors $K$ correlates with degree. This will have an effect similar to preferential attachment, that high degree nodes are more likely to gain new edges. This effect may give rise to an almost-power law degree distribution.

## 4.2 K process

We propose a model for the evolution of common neighbors matrix $K$ independent from the graph. Here, we will disassociate the common neighbors values from the graph, and index using $i$ and $j$ as opposed to $v$ and $w$. This is to reinforce the idea that, under this model for $K$, there is no underlying graph, and that $K_{ij}(t)$ is simply a quantity that evolves on its own.

We will also use the notation

$$N_i(t) = \sum_j K_{ij}(t)$$

to denote the total number of common neighbors of node $i$ at time $t$.

We now propose a decoupled model for $K_{ij}(t)$, which we will call the $K$ process. When a new node $k$ enters the system, we set $K_{k\ell} = 0$ for all $\ell$ in the vertex set. At time $t+1$, select $i$ uniformly from the node set. Select $\eta(t) = \lfloor cn(t)^\theta \rfloor$ nodes $j_1, \ldots, j_{\eta(t)}$ proportional to $K_{ij_m}(t) + \delta$ for some small constant $\delta$, and some constant $c$. Then update

$$K_{ij_m}(t+1) = K_{ij_m}(t) + 1, \quad m = 1, \ldots, \eta(t). \tag{4.1}$$

The rational behind this model is as follows. In the graph model, at time $t+1$, some vertex $i$ sends an invitation and perhaps forms a new connection. When $i$ forms a new connection (say, to $k$), some number of other nodes $(j_1, \ldots, j_m)$ in the graph will now have an additional common neighbor with $i$. Suppose that with each new connection, about $n(t)^\theta$ nodes have a new common neighbor with $i$ (and others now have a new common neighbor with $k$). We assume $0 \le \theta \le 1$. We assume that the selected $j$'s tend to be "close" to $i$ in some sense, perhaps

with many common neighbors (that is, $K_{ij}(t)$ is relatively high for selected vertices $j$).

For our partial results, we consider a fixed node $i$ and consider the distribution the $K_{ij}(t)$. In Section 4.2.1, we give an asymptotic result on $N_i(t)$, and in Section 4.2.2, we consider $(K_{ij}(t), j \geq 0)$ for large $t$.

### 4.2.1 Total Common Neighbors

Define $N_i(t) = \sum_j K_{ij}(t)$ to be the total number of common neighbors between $i$ and other nodes.

**Theorem 17.** *Under the K process, there exists a random variable $Z_\theta$ such that for any fixed $i$*

$$\lim_{t\to\infty} \frac{N_i(t)}{t^\theta} = Z(\theta)$$

*in probability.*

We will prove this in two parts. First, a coupling argument in Lemma 10, and then find the limiting distribution of the coupled quantity in Lemma 11.

Define a "success" at time $t$ to be if node $i$ is chosen first in the common neighbor process above. Let $\mathbb{1}_t$ be 1 if a success happens at time $t$, and 0 otherwise. A success happens with probability $(1-\alpha)/n(t)$. For each success, $N_i(t)$ increases by $\eta(t) = \lfloor cn(t)^\theta \rfloor$. We write

$$N_i(t) = \sum_{s=1}^{t-1} \mathbb{1}_s \eta(s). \tag{4.2}$$

The expectation is

$$\begin{aligned} EN_i(t) &= \sum_{s=1}^{t-1} E\left[E[\mathbb{1}_s \eta(s)|n(s)]\right] \\ &= \sum_{s=1}^{t-1} E\left[(1-\alpha)\frac{\eta(s)}{n(s)}\right] \\ &= (1-\alpha)\sum_{s=1}^{t-1} E\left[\frac{\eta(s)}{n(s)}\right]. \end{aligned}$$

$N_i(t)$ has a lot of dependent parts, so we hope to say it is close to something that is easier to work with. Let $U_s \sim U(0,1)$, $s \geq 1$, be iid, and let

$$\chi_s = \mathbb{1}\{U_s \leq \frac{1-\alpha}{n(s)}\}, \qquad \hat{\chi}_s = \mathbb{1}\{U_s \leq \frac{1-\alpha}{\alpha s}\},$$

$$N(t) = \sum_{s=1}^{t} \eta(s)\chi_s, \qquad \hat{N}(t) = \sum_{s=1}^{t} c(\alpha s)^{\theta}\hat{\chi}_s \tag{4.3}$$

**Lemma 10.** *Fix $i$ and let $N(t) = N_i(t)$. For $\hat{N}(t)$ defined in 4.3,*

$$\lim_{t\to\infty} \frac{N(t)-\hat{N}(t)}{t^{\theta}} = 0$$

*in probability.*

*Proof.* It is enough to show that the expectation of the absolute value converges to 0.

$$E\left|\frac{N(t)-\hat{N}(t)}{t^{\theta}}\right| \leq t^{-\theta}\sum_{s=1}^{t} E\left|\eta(s)\chi_s - c(\alpha s)^{\theta}\hat{\chi}_s\right|.$$

Adding and subtracting $\eta(s)\hat{\chi}_s$,

$$E\left|\frac{N(t)-\hat{N}(t)}{t^{\theta}}\right| \leq t^{-\theta}\sum_{s=1}^{t}\left(E\left|\eta(s)(\chi_s-\hat{\chi}_s)\right| + E\left|\hat{\chi}_s(\eta(s)-c(\alpha s)^{\theta})\right|\right).$$

Recall that $\chi'_s = 1$ with probability $(1-\alpha)/\alpha s$ and that $|\chi_s - \chi'_s| = 1$ with probability, conditional on $n(s)$, of $(1-\alpha)|1/n(s) - 1/\alpha s|$. Otherwise $|\chi_s - \chi'_s| = 0$. So

$$E\left|\frac{N(t)-\hat{N}(t)}{t^{\theta}}\right| \leq (1-\alpha)t^{-\theta}\sum_{s=1}^{t}\left(E\left[\eta(s)\left|\frac{1}{n(s)}-\frac{1}{\alpha s}\right|\right] + \frac{1}{\alpha s}E\left|\eta(s)-c(\alpha s)^{\theta}\right|\right). \tag{4.4}$$

Will prove the inside of the above sum is $o(s^{\theta-1})$, so that the entire RHS approaches 0 as $t \to \infty$.

Consider

$$\frac{E|\eta(s) - c(\alpha s)^\theta|}{s^\theta} = E\left|\frac{\lfloor cn(s)^\theta \rfloor}{s^\theta} - c\alpha^\theta\right|.$$

Define

$$X_s = \frac{\lfloor cn(s)^\theta \rfloor}{s^\theta} - c\alpha^\theta. \tag{4.5}$$

We will show $E|X_s| \to 0$ as $s \to \infty$.

By LLN, $\frac{n(s)}{s} \to \alpha$ a.s., so with probability 1, $X_s \to 0$. If the variance of $X_s$ is bounded for all $s$, we can then say that the expectation of its absolute value converges to 0. Using the fact that $n(s) \leq s$, and $(a-b)^2 \leq 2a^2 + 2b^2$ for real numbers $a$ and $b$,

$$EX_s^2 = E\left[\left(\frac{\lfloor cn(s)^\theta \rfloor}{s^\theta} - c\alpha^\theta\right)^2\right] \leq 2E\left|\left(\frac{\lfloor cn(s)^\theta \rfloor}{s^\theta}\right)^2\right| + 2c^2\alpha^{2\theta} \leq 2c^2(1+\alpha^{2\theta}) < \infty.$$

Hence the variance is finite and $E|X_s| \to 0$. This implies that $E|\eta(s) - c(\alpha s)^\theta| = o(s^\theta)$.

Secondly, define

$$Y_s = \frac{\eta(s)}{n(s)s^{\theta-1}} - \frac{\eta(s)}{\alpha s^\theta} \tag{4.6}$$

and consider

$$\frac{1}{s^{\theta-1}} E\left[\eta(s)\left|\frac{1}{n(s)} - \frac{1}{\alpha s}\right|\right] = E|Y_s|.$$

We will prove that $E|Y_s| \to 0$. Again, we know that $n(s)/s \to \alpha$ almost surely as

$s \to \infty$, so $Y_s \to 0$ almost surely. Further,

$$\begin{aligned} EY_s^2 &= E\left[\left(\frac{\eta(s)}{n(s)s^{\theta-1}} - \frac{\eta(s)}{\alpha s^\theta}\right)^2\right] \\ &\leq E\left[2\left(\frac{\eta(s)}{n(s)s^{\theta-1}}\right)^2 + 2\left(\frac{\eta(s)}{\alpha s^\theta}\right)^2\right] \\ &\leq 2E\left[\frac{c^2 n(s)^{2(\theta-1)}}{s^{2(\theta-1)}} + \frac{c^2 n(s)^{2\theta}}{\alpha^2 s^{2\theta}}\right] \\ &\leq 2c^2\left(1 + \frac{1}{\alpha^2}\right) < \infty. \end{aligned}$$

Hence the variance of $Y_s$ is finite for all $s$ and we can say $E|Y_s| \to 0$. This implies that $E[\eta(s)|\frac{1}{n(s)} - \frac{1}{\alpha s}| = o(s^{\theta-1})$.

Returning to (4.4), we have

$$\begin{aligned} E\left|\frac{N(t) - \hat{N}(t)}{t^\theta}\right| &\leq (1-\alpha)t^{-\theta}\sum_{s=1}^{t}\left(o(s^{\theta-1}) + \frac{1}{\alpha s}o(s^\theta)\right) \\ &= (1-\alpha)t^{-\theta}\sum_{s=1}^{t} o(s^{\theta-1}) \qquad (4.7) \\ &= (1-\alpha)t^{-\theta}\cdot o(t^\theta) \\ &= o(1). \end{aligned}$$

□

Next step is to prove $\widehat{N}(t)/t^\theta$ converges to some random variable as $t \to \infty$.

**Lemma 11.** *There exists a random variable $Z(\theta)$ such that*

$$\lim_{t\to\infty}\frac{\hat{N}(t)}{t^\theta} = Z(\theta).$$

*Proof.* Taking Laplace transform,

$$\begin{aligned} E\left[e^{-ct^{-\theta}\widehat{N}(t)}\right] &= \prod_{s=1}^{t} E\left[e^{-ct^{-\theta}(\alpha s)^{\theta}\chi'_s}\right] \\ &= \prod_{s=1}^{t}\left[1-\frac{1-\alpha}{\alpha s}\left(1-e^{-ct^{-\theta}(\alpha s)^{\theta}}\right)\right] \\ &= \exp\left\{\sum_{s=1}^{t}\log\left[1-\frac{1-\alpha}{\alpha s}\left(1-e^{-ct^{-\theta}(\alpha s)^{\theta}}\right)\right]\right\}. \end{aligned} \tag{4.8}$$

Let $M$ be large enough that for $s \geq M$, $\frac{1-\alpha}{\alpha s}(1-e^{-c\alpha^{\theta}}) < 1$.

Split the sum in (4.8) as follows:

$$\begin{aligned} E\left[e^{-ct^{-\theta}\widehat{N}(t)}\right] = \exp\Bigg\{&\sum_{s=1}^{M-1}\log\left[1-\frac{1-\alpha}{\alpha s}\left(1-e^{-ct^{-\theta}(\alpha s)^{\theta}}\right)\right] \\ &+\sum_{s=M}^{t}\log\left[1-\frac{1-\alpha}{\alpha s}\left(1-e^{-ct^{-\theta}(\alpha s)^{\theta}}\right)\right]\Bigg\}. \end{aligned} \tag{4.9}$$

Consider the first sum. For fixed $s$ and as $t \to \infty$, $e^{-c(\alpha s/t)^{\theta}} \to 1$.

Then

$$\lim_{t\to\infty}\sum_{s=1}^{M-1}\log\left[1-\frac{1-\alpha}{\alpha s}\left(1-e^{-ct^{-\theta}(\alpha s)^{\theta}}\right)\right] = 0. \tag{4.10}$$

Let $x = \frac{1-\alpha}{\alpha s}(1-e^{-c(\alpha s/t)^{\theta}})$. For $s \geq M$, $x \leq \frac{1-\alpha}{\alpha s}(1-e^{-c\alpha^{\theta}}) < 1$, so we use the approximation $\log(1-x) = -x + O(x^2)$.

The second sum of (4.9) becomes

$$\sum_{s=M}^{t}\left[-\frac{1-\alpha}{\alpha s}\left(1-e^{-c(\alpha s/t)^{\theta}}\right)+O(x^2)\right]. \tag{4.11}$$

Since $x \leq \frac{1-\alpha}{\alpha s}$, $O(x^2) = O(s^{-2})$. Now we have

$$-\frac{1-\alpha}{\alpha}\sum_{s=M}^{t}\frac{1}{s}(1-e^{-c(\alpha s/t)^{\theta}}) + \sum_{s=M}^{t}O(s^{-2}) \tag{4.12}$$

There error term sums to $O(t^{-1})$, so the error goes to 0 in the limit. Consider $\sum_{s=M}^{t} \frac{1}{s}(1 - e^{-c(\alpha s/t)^\theta})$. Multiplying and dividing by $t$, this becomes a Riemann sum. Taking $t \to \infty$, we have

$$\begin{aligned}\lim_{t\to\infty} \sum_{s=1}^{t} \frac{1}{t} \cdot \frac{t}{s}\left(1 - e^{-c(\alpha s/t)^\theta}\right) &= \int_0^1 \frac{1}{x}(1 - e^{-c\alpha^\theta x^\theta})dx \\ &= \frac{1}{\theta}\int_0^1 \frac{1}{y}\left(1 - e^{-c\alpha^\theta y}\right)dy,\end{aligned}$$

which looks like an ID law. Overall, we have

$$\lim_{t\to\infty} E\left[e^{-ct^{-\theta}\widehat{N}(t)}\right] = \exp\left\{-\frac{1-\alpha}{\alpha\theta}\int_0^1 \frac{1}{y}(1 - e^{-c\alpha^\theta y})dy\right\}. \tag{4.13}$$

To identify the limiting distribution of $\hat{N}(t)/t^\theta$, we notice that (4.13) fits the Lévy Khintchine representation. For $X$ a nonnegative ID random variable,

$$Ee^{-\phi X} = \exp\left\{-\frac{1}{2}\phi^2\sigma^2 + \int_0^\infty \left(e^{-\phi x} - 1\right)\mu(dx) - \phi b\right\} \tag{4.14}$$

for $\phi \in \mathbb{R}$. First let's rewrite (4.13) with change of variables $\alpha^\theta y \mapsto z$. This yields

$$\exp\left\{\frac{1-\alpha}{\alpha\theta}\int_0^{\alpha^\theta} \frac{1}{z}(e^{-\phi z} - 1)dz\right\}. \tag{4.15}$$

Matching (4.15) to (4.14), we have $\sigma^2 = 0$, $b = 0$, and

$$\mu(dx) = \frac{1-\alpha}{\alpha\theta}\frac{1}{x}dx, \qquad 0 < x \le \alpha^\theta \tag{4.16}$$

and 0 elsewhere.

To find the density of this distribution, we take the inverse Fourier transform of the characteristic function. First we need to check that the integral in the characteristic function

$$\phi(t) = \exp\left\{\frac{1-\alpha}{\alpha\theta}\int_0^{\alpha^\theta} \frac{1}{z}\left(e^{itz} - 1\right)dz\right\} \tag{4.17}$$

is finite. We'll consider the real and imaginary parts separately. For the real part:

$$\int_0^{\alpha^\theta} \frac{\cos(\phi z) - 1}{z} dz \tag{4.18}$$

note $\lim_{z\to 0} \frac{\cos(\phi z)-1}{z} = 0$, and the integrand is continuous and finite on the rest of the interval $(0, \alpha^\theta)$.

For the imaginary part:

$$i \int_0^{\alpha^\theta} \frac{\sin(\phi z)}{z} dz \tag{4.19}$$

note $\lim_{z\to 0} \frac{\sin(\phi z)}{z} = \phi$, and the integrand is continuous and finite on the rest of the interval $(0, \alpha^\theta)$. Hence the whole thing is integrable.

Now we need to check integrability of the characteristic function, that is, we need

$$\int_{-\infty}^{\infty} |\phi(t)| dt < \infty.$$

Suppose

$$\phi(t) = \exp\{x(t) + iy(t)\}$$

Then

$$|\phi(t)|^2 = \exp\{x(t) + iy(t\} \exp\{x(t) - iy(t)\} = \left(e^{x(t)}\right)^2.$$

So

$$\int_{-\infty}^{\infty} |\phi(t)| dt = \int_{-\infty}^{\infty} e^{x(t)} dt. \tag{4.20}$$

Recall $x(t)$ is simply the real part of the integral in (4.17), so that

$$x(t) = -\frac{1-\alpha}{\alpha\theta} \int_0^{\alpha^\theta t} \frac{1 - \cos(z)}{z} dz.$$

This is the well-known cosine integral

$$Cin(x) = \int_0^x \frac{1 - \cos t}{t} dt.$$

It is related to

$$Ci(x) = -\int_x^\infty \frac{\cos t}{t} dt$$

by

$$Cin(x) = \gamma + \ln x - Ci(x),$$

where $\gamma$ is the Euler-Mascheroni constant. It is known that $Ci(x) < 1$ for all $x \geq 0$ (see, e.g., [84]). Therefore

$$Cin(x) \geq \gamma + \ln x - 1$$

for all $x \geq 0$. For our $t \geq 0$, then

$$x(t) \leq -\frac{1-\alpha}{\alpha\theta}\left(\gamma - 1 + \ln(\alpha^\theta t)\right).$$

Since $Cin(x)$ is an even function, $Cin(-x) = Cin(x)$. Therefore even when $t < 0$,

$$\begin{aligned} x(t) &= \frac{1-\alpha}{\alpha\theta}\int_{\alpha^\theta t}^0 \frac{1-\cos z}{z} dz \\ &= \frac{1-\alpha}{\alpha\theta}\int_{\alpha^\theta t}^0 -\frac{1-\cos|z|}{|z|} dz \\ &= -\frac{1-\alpha}{\alpha\theta}\int_0^{|\alpha^\theta t|} \frac{1-\cos(z)}{z} dz \\ &\leq -\frac{1-\alpha}{\alpha\theta}\left(\gamma - 1 + \ln(|\alpha^\theta t|)\right). \end{aligned}$$

Now,

$$\int_{-\infty}^\infty e^{x(t)} dt = e^{-(1-\alpha)(\gamma-1)/(\alpha\theta)} \int_{-\infty}^\infty |\alpha^\theta t|^{-\frac{1-\alpha}{\alpha\theta}} dt$$

which converges if

$$\frac{1-\alpha}{\alpha\theta} > 1.$$

□

Let $Z(\theta)$ be the random variable with characteristic function given by (4.17). Lemmas 10 and 11 are sufficient to show that as $t \to \infty$,

$$\frac{N(t)}{t^\theta} \to Z(\theta)$$

in probability. So for large $t$, the total number of common neighbors $N_i(t)$ of node $i$, is approximately

$$N_i(t) \approx t^\theta Z_i(\theta).$$

### 4.2.2 Common Neighbors between $i$ and $j$

We now know that $N_i(t)/t^\theta$ converges to some random variable as $t \to \infty$. It remains to understand $K_{ij}(t)$, the number of common neighbors between nodes $i$ and $j$ at time $t$.

In our common neighbor process, we choose a node $i$ uniformly at random to send out an invitation. Then $\eta(t) = \lfloor cn(t)^\theta \rfloor$ nodes $j$ are chosen to increase $K_{ij}(t)$. Fixing some node $j$, the probability of increasing $K_{ij}(t)$ given that $i$ was selected is

$$P(j \text{ chosen}|i) = \frac{K_{ij}(t)}{N_i(t)}.$$

Fix $i$ for the rest of the section. We are interested in the distribution of $\Big(K_{ij}(t), j \geq 1\Big)$. Let

$$Y_k(t) = |\{1 \leq j \leq n(t) : K_{ij}(t) = k\}| \tag{4.21}$$

be the number of nodes $j$ such that $i$ and $j$ share $k$ common neighbors at time $t$. Let $P_k(t) = \frac{Y_k(t)}{n(t)}$ be the proportion of nodes that share $k$ common neighbors with $i$ at time $t$.

We now give a heuristic argument to understand the evolution of $E[Y_k(t)]$ as $t \to \infty$. The details are left out, as the approach here is similar to many derivations of the degree distribution in a preferential attachment model (see [39]).

$Y_k(t)$'s evolve according to:

$$E[Y_k(t+1) - Y_k(t) \mid G(t)] = \frac{1}{n(t)} \frac{cn(t)^\theta}{N_i(t)} [(k-1)Y_{k-1}(t) - kY_k(t)] + \alpha \mathbb{1}_{k=0}.$$

Using $n(t) \approx \alpha t$ and $N_i(t) \approx t^\theta Z_i$,

$$E[Y_k(t+1) - Y_k(t) \mid G(t), Z_i] \approx \frac{c\alpha^{\theta-1}}{Z_i t} [(k-1)Y_{k-1}(t) - kY_k(t)] + \alpha \mathbb{1}_{k=0}.$$

Let $A = c\alpha^{\theta-1} E[Z_i^{-1}]$. Then taking an additional expectation yields

$$E[Y_k(t+1)] \approx E[Y_k(t)] \left[1 - \frac{A}{t} k\right] + \frac{A}{t}(k-1)E[Y_{k-1}(t)] \tag{4.22}$$

for $k > 0$, and

$$E[Y_0(t+1)] = E[Y_0(t)] + \alpha = \alpha t. \tag{4.23}$$

for $k = 0$. Assuming that $Y_k(t)/t \to p_k$ as $t \to \infty$ for all $k \geq 0$, then (4.22) and (4.23) imply that

$$p_0 = \alpha$$

and

$$p_k = A(k-1)p_{k-1} - Akp_k \tag{4.24}$$

for $k \geq 1$.

$$p_k = \left(\frac{A(k-1)}{1+Ak}\right) p_{k-1}, \qquad k \geq 1, \tag{4.25}$$

where $A$ is the random variable $c\alpha^{\theta-1} Z_i^{-1}$.

## 4.3 Common Neighbors and Jaccard Similarity

Our goal in this section is to understand the distribution of the number of common neighbors in a graph. We consider the average number of common neighbors, and compare this to node degree. This will give us a better sense of how degree and common neighbors distribution are related. We also consider Jaccard similarity in relation to degree, and in relation to common neighbors. Surprisingly, we see that for nodes with certain local structures there is an inverse relationship between average common neighbors and Jaccard similarity.

We simulated CNM with $\alpha = 0.2$ and $\beta = 0.1$ for 100,000 time steps, to end up with about 20,000 nodes. For each node $v$, we find the number of common neighbors $K_{vu}$ for every other node $u$. This value is averaged over all non-negative values, to get

$$K_{avg} = \frac{\sum_u K_{vu}}{\sum_u \mathbb{1}(K_{vu} > 0)}.$$

In Figure 4.1(b), we plot average number of common neighbors against degree, for the simulated graph. We also find an average Jaccard similarity $J$, and plot it against degree in Figure 4.1(a).

We observe three types of nodes, according to their local connectivity structure. These three types correspond to certain regions of the plots, as identified in Figure 4.2. We counterintuitively see high $K$ and $J$ values for nodes of small degree and low $K$ and $J$ values for nodes of high degree, although there may be large variance in these quantities for high-degree nodes. We observed similar results in real-world network data.

Type (1): Nodes of small degree with a well-connected group of neighbors,

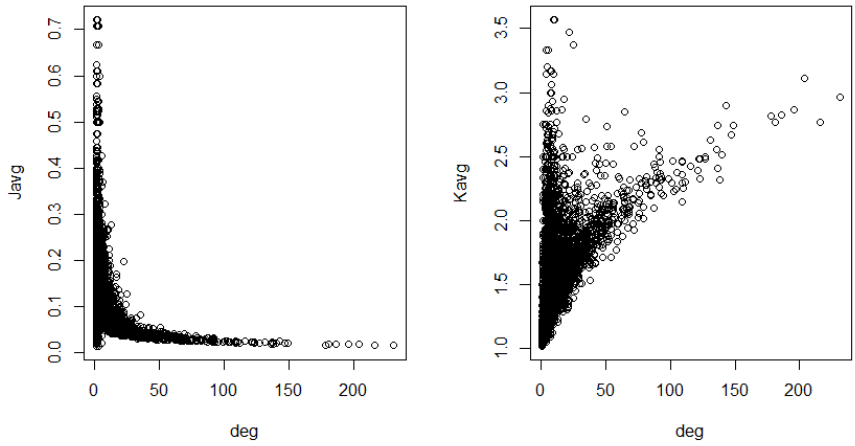



Figure 4.1: In the simulated CNM, (a) average Jaccard similarity and (b) average common neighbors $K$ are plotted against degree.

almost forming a clique, as in Figure 4.3. Nodes of this type share a large number of common neighbors with these few nodes, but in general have very few neighbors in common with the other nodes. Thus their $K$ distribution is lumped into two categories: either surprisingly high, or zero. These nodes also tend to have high Jaccard similarity ($J$ close to 1) with their local nodes, as the number of common neighbors is close to total degree.

A surprising subset of this category are nodes of small degree, which share the aforementioned results, even though their local structure is not clique-like. See Figure 4.4(b), where node 1 has high Jaccard similarity $J$ and a high number of common neighbors with node 7. Even in Figure 4.4(a), node 1 has a high Jaccard similarity with node 3, even though it has very few connections. These types of peripheral cases should be taken into account when considering a model like CNM. When a user decides whether to use number of common neighbors or Jaccard similarity in the fitness function, they should consider whether the wish special cases like these to have an advantage or disadvantage.

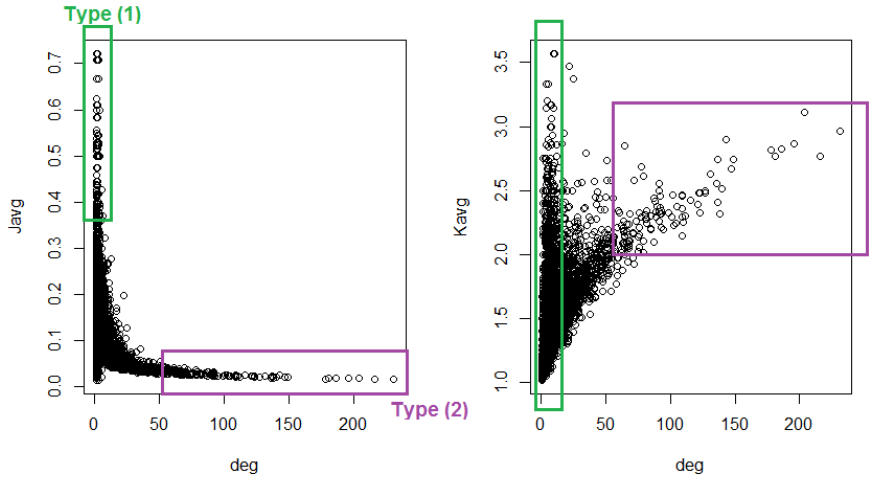



Figure 4.2: In the simulated CNM, (a) average Jaccard similarity and (b) average common neighbors $K$ are plotted against degree. Nodes of Type (1) are outlined in green, Type (2) outlined in purple, and Type (3) the remaining.

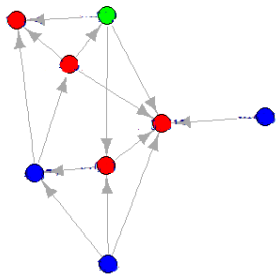


Figure 4.3: The neighborhood and friends-of-friends of a particular node (green) with low total degree but relatively high average number of common neighbors.

Type (2): Nodes with very high degree (locally star-like). These nodes have more of a gradient in their common neighbor distribution, but on average, they tend to have relatively few neighbors in common with other nodes. The average number of common neighbors is subject to an averaging effect. They may have many common neighbors with one or two other nodes, but for most other nodes, they only have a few neighbors in common. Their Jaccard similarity

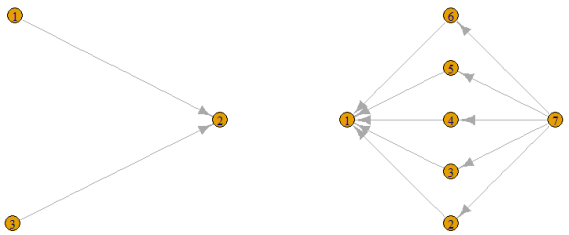


Figure 4.4: Illustration of special cases: (a) a graph of 3 nodes. (b) a graph of 7 nodes, where node 1 has 5 common neighbors with node 7.

tends to be surprisingly low, due no doubt to their high degree.

Type (3): Nodes with neither low or high degree, neither locally clique-like or locally star-like. These nodes tend to be have $K$ and $J$ distributions that look like a mixture of those for the above two types. We conjecture that the "V" shape in Figure 4.1 is due to this averaging effect. As degree increases, there are fewer and fewer zero values of $K$, but also fewer nodes with which they have high Jaccard similarity.

Thus we have counterintuitively seen high $K$ and $J$ values for nodes of small degree and low $K$ and $J$ values for nodes of high degree, although there may be large variance in these quantities for high-degree nodes.

.

## 4.4 Modularity and Cluster Coefficient

The goal of this section is to compare graph properties of the CNM to real-world networks. The properties of interest are modularity, cluster coefficient, etc. We summarize the data in Tables 4.1 and 4.2. The real network data come from a collaboration network and a citation network.

- **Collaboration network**: Data from Stanford SNAP collection, papers submitted to the Arxiv category *General Relativity and Quantum Cosmology* between January 1993 to April 2003 [56]. If author $i$ coauthored a paper with author $j$, then the graph contains an undirected edge $(i, j)$.

  The graph contains 5242 nodes and 14496 edges.

- **Citation Network**: High energy physics theory citation graph is from the e-print arXiv and covers all the citations within a dataset of 27,770 papers with 352,807 edges [55, 34]. If a paper i cites paper j, the graph contains a directed edge from i to j.

  We extract a subgraph of 12,000 nodes and 68,861 edges.

We simulate the CNM as defined in Section 4.1, as well as the version using Jaccard similarity. We also simulate two versions of preferential attachment models for reference. Given a number of nodes and a number of edges, the simulated graphs are generated as described below.

- **Common Neighbors simulation**: Let $G_0$ have one node and zero edges. Form $G_{t+1}$ by adding onto $G_t$:

  1. Add a new node with probability $\alpha = \frac{\text{number of nodes}}{\text{number of edges}}$.

2. Select node $v$ uniformly at random.
3. With probability $\beta = 0.1$, select $w$ uniformly at random.
4. Otherwise, calculate the number of common neighbors $K_t(v, \cdot)$ between $v$ and every other node, and choose $w$ with probability proportional to $K_t(v, w)$.
5. Form edge $(v, w)$.

Repeat for [number of edges] time steps.

- **Jaccard simulation**: Let $G_0$ have one node and zero edges. Form $G_{t+1}$ by adding onto $G_t$:

1. Add a new node with probability $\alpha = \frac{\text{number of nodes}}{\text{number of edges}}$.
2. Select node $v$ uniformly at random.
3. With probability $\beta = 0.1$, select $w$ uniformly at random.
4. Otherwise, calculate the Jaccard similarity $J_t(v, \cdot)$ between $v$ and every other node, and choose $w$ with probability proportional to $J_t(v, w)$.

$$J_t(v, w) = \frac{|N_v \cap N_w|}{|N_v \cup N_w|}.$$

5. Form edge $(v, w)$.

Repeat for [number of edges] time steps.

- **Bollobás PAM simulation**: Let $G_0$ have one node and zero edges. Form $G_{t+1}$ by adding onto $G_t$:

1. Add a new node $v$ with probability $\alpha = \frac{\text{number of nodes}}{\text{number of edges}}$. Otherwise select $v$ with probability proportional to its out-degree plus $\delta_{out} = 1$.
2. Select $w$ with probability proportional to its in-degree plus $\delta_{in} = 1$.

3. Form edge $(v, w)$.

Repeat for [number of edges] time steps.

- **Barabási-Albert PAM simulation** Let $m = \frac{\text{number of edges}}{\text{number of nodes}}$. (Note: For the collaboration graph trials, we used $m = 1$.) Let $G_0$ have $m$ nodes and zero edges. Form $G_{t+1}$ by adding onto $G_t$:

  1. Add a new node, $v$, with $m$out-edges.
  2. Select nodes $w_1, \ldots, w_m$ with probability proportional to its total degree plus 1.
  3. Form edges $(v, w_i)$.

  Repeat for [number of nodes] time steps.

For each graph, we found the following quantities, which are summarized in Tables 4.1 and 4.2.

- WT: the modularity calculated from dividing the graph into communities via the Walktrap algorithm [73].(Recall the definition of modularity (1.1) on page 6.)
- FG: the modularity calculated from dividing the graph into communities via the Fast-Greedy algorithm [21]
- GCC: the global clustering coefficient of the entire graph, i.e., $\frac{\text{number of triangles}}{\text{number of triples}}$.
- ACC: the average clustering coefficient. Find local CC for all nodes, and then average. Local CC is the number of triangles connected to the vertex, divided by the number of triples centered on the vertex.

| Graph | WT | FG | GCC | ACC | Trials |
|---|---|---|---|---|---|
| Collaboration | 0.78 | 0.80 | 0.63 | 0.69 | N/A |
| Common Neighbors | 0.73 | 0.70 | 0.22 | | 32 |
| Jaccard | 0.66 | 0.67 | 0.25 | 0.49 | 28 |
| Bollobás | 0.17 | 0.41 | 0.02 | 0.13 | 100 |
| Barabási-Albert | 0.89 | 0.98 | 0 | 0 | 100 |

Table 4.1: Data for Collaboration graph. All graphs have 5,000 nodes and 15,000 edges, except for the Barabási-Albert model, which has 5,000 nodes and 5,000 edges.

| Graph | WT | FG | GCC | ACC | Diam | Path | Giant | Trials |
|---|---|---|---|---|---|---|---|---|
| Citation | 0.57 | 0.51 | 0.12 | 0.33 | 13 | 4.3 | 0.99 | N/A |
| Citation Subgraph | 0.57 | 0.55 | 0.12 | 0.32 | 17 | 4.9 | 0.90 | N/A |
| Common Neighbors | 0.34 | 0.38 | 0.096 | 0.34 | 16 | 4.4 | 0.81 | 10 |
| Jaccard | 0.41 | 0.45 | 0.14 | 0.39 | 15 | 4.7 | 0.86 | 1 |
| PA Bollobás | 0.079 | 0.28 | 0.011 | 0.35 | 7.9 | 2.6 | 1 | 10 |
| PA Barabási-Albert | 0.13 | 0.27 | 0.0038 | 0.14 | 5 | 2.8 | 1 | 10 |

Table 4.2: Data for Citation graph. Citation graph size: 28,000 nodes and 350,000 edges. Subgraph size and simulation size: 12,000 nodes and 68,861 edges.

- Diam: the length of the longest shortest path
- Path: the average of the shortest path lengths over all node pairs
- Giant: proportion of nodes in the largest connected component. If 1, then graph is connected
- Trials: For the 4 types of simulated graphs, we ran multiple trials and average the above quantities over all trials.

Our simulations are admittedly limited, but we see that the CNM and Jaccard model in general out perform the PA models in terms of modularity and cluster coefficient. In Table 4.1, the Barabási-Albert model considered is a preferential attachment tree, which gives it high modularity and 0 cluster coefficient. In reference to Table 4.2 for the citation network, we see that the citation graph actually has much less of a small-world property than results from a preferential

attachment model. The average path length and longest path length (diameter) are both much higher than the corresponding quantities in the PA graphs. Encouragingly, both the CNM and the Jaccard model have similar path values to the citation graph.

We further note that these observations are not the result of fitting numerous parameters. The decision parameter $\beta = 0.1$ was chosen arbitrarily. The only fitted parameter was node arrival rate, $\alpha$, which was chosen, not to optimize modularity or cluster coefficient, but solely to create graphs of the same size as the real-world graphs.

## 4.5 Conclusion

In this chapter, we defined the Common Neighbors Model based on the assumption that attachments is made with preference to node with which a high number of common neighbors are shared. Although we found this model to be intractable, we have shown computational evidence that this model generates graph that have similar community and clustering to read-world citation and collaboration graphs. This model has the advantage of creating communities organically, without reliance on node labels and a matrix parameter.

CHAPTER 5

# INFORMATION DIFFUSION ON NETWORKS

## 5.1 Introduction

The study of epidemic spreading processes on complex networks has long been a research topic of interest. Traditionally and terminologically, such spreading processes describe the transmission of an infectious disease via interactions among a population, where potential interactions are characterized by the network. The research is equally suited to other spreading processes, such as the dissemination of information or novel behaviors. There are a range of models for stochastic epidemic-like behavior, including the ubiquitous SIR and SIS models, and numerous variants [68].

These stochastic models are well-understood for certain graph structures - particularly for regular graphs, in which all nodes have the same degree. However, in more complicated structures, such as graphs arising from the preferential attachment model, there is not as strong an understanding. This is not due to a lack of research in the area, but rather the complexity and intractability of the problem. Some studies take a probabilistic approach to studying the exact solution [71, 80, 48], but the majority of works rely on deterministic approximations and simulations [68].

The aim of this chapter is to provide new insights to two specific open questions. Our first area of contribution is in the class of networks with community structure. Recent work on the existence of tightly-knit groups in certain networks, and the huge field of community detection, emphasize the impor-

tance of understanding networks with this structure (see Section 1.1 for more detail). Other authors have published simulation studies considering the impact of community structure on spreading processes, but with mixed results. In Section 5.2, we give an overview of the previous work and detail our contributions, giving explicit discussion on what we can and cannot hope to learn from these types of experiments. Our main result proves that the so-called epidemic threshold for a general network is slightly lower than, but altogether comparable to, the epidemic threshold of its dominant community. This result supports our claim that the existence of "community structure" in a network is less important in understanding the relevant spreading processes than the actual structure of the individual communities and connecting edges.

The second area of contribution is on time-varying or temporal networks, in which nodes and edges exist for only a subset of the entire time scale [26, 62, 85]. These types of networks are of great interest in general, as they have a more general range of applications, including networks in which time of day/week/year are of importance in the connectivity structure. Our work focuses solely on the class of networks that grow over time. In Section 5.3, we provide a detailed simulation study of the lifetime of spreading processes on a growing graph compared to a static (non-growing) graph, and we give evidence to support the possibility of infinite survival of spreading processes on this type of network.

All of our work focuses on the SIS model, and should not be expected to transfer ad verbum to SIR or other types of models, although some intuition may be gained.

Under the SIS model, nodes exist in one of two possible states: susceptible and infected, and may change state over time. Infected nodes may transmit

the infection to susceptible nodes via edge interactions, and with the passage of time, infected nodes eventually "recover" from the disease, reverting to susceptibility. Note that unlike the SIR model, for example, there is no immunity to the infection, meaning that a node may contract the infection multiple times over the lifetime of the disease. Under this formulation the infection may persist in the network for an arbitrarily long amount of time.

In Section 5.1.1, we detail the SIS model and previous work, before discussing the new contributions in Sections 5.2 and 5.3.

### 5.1.1 SIS Model

A continuous-time SIS model on a finite graph $G$ is characterized by an infection rate $\nu > 0$ and a recovery rate $\delta > 0$. On a graph of $n$ nodes, we define the vector process $(X_t, t \geq 0)$, $X_t \in \{0,1\}^n$, where we say that a node $i$ is infected at time $t$ if $X_t(i) = 1$, and susceptible if $X_t(i) = 0$.

An infected node becomes susceptible again after an exponentially distributed amount of time with rate $\delta$. For every edge $(i, j)$, there is a Poisson process with rate $\nu$ representing contact between nodes $i$ and $j$. If one node is susceptible and the other is infected at a contact time, then the susceptible node becomes infected. That is, if there is an arrival at time $t$, and $X_{t^-}(i) = 1$, $X_{t^-}(j) = 0$, then we set $X_t(j) = 1$.

We often consider, without loss of generality, the effective spreading rate $\beta = \nu/\delta$, and consider the recovery rate to be 1. Thus the parameter space is truly only one-dimensional, in $\beta > 0$ [69].

Often we are interested in a critical spreading rate $\beta_c$, above which there is a positive probability that an infection will survive for a sufficiently long time. On an infinite graph $G_\infty$, a process with spreading rate $\beta > \beta_c$ has positive probability of infinite survival, and a process with spreading rate $\beta < \beta_c$ is guaranteed to die out in finite time. It has been stated that infinite scale-free networks have $\beta_c = 0$ [9].

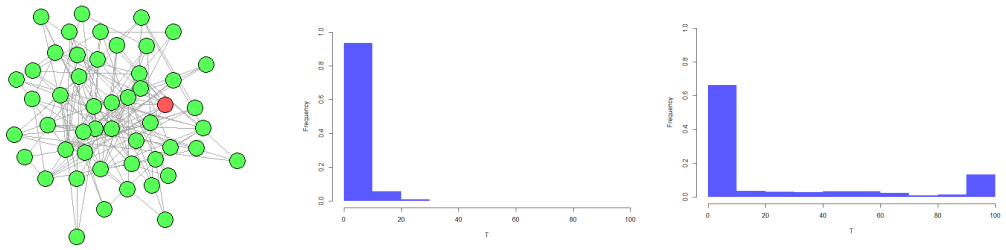
Finite graphs necessarily have finite lifetime, so [33] divides extinction time into two regimes. On a finite graph with $n$ nodes, extinction is said to be quick when $E[\tau] = O(\log n)$ and slow when $E[\tau] = \Omega(e^{n^\alpha})$ for some $\alpha > 0$. It has been established that a sufficient condition for quick extinction on finite graphs is $\beta < \frac{1}{\rho}$, where $\rho$ is the spectral radius of the graph [33]. Then there must exist a critical value $\beta_c \geq \frac{1}{\rho}$, above which the process may survive for a long amount of time.

For finite graphs grown via preferential attachment, the critical value is shown to be inversely related to the size of the graph [33]. As a preferential attachment graph increases in size, the critical value decreases, with limit 0 as size increases to infinity.

For simulation, we consider a discrete-time model where, instead of a spreading rate $\beta$, we have a spreading probability

$$\lambda = 1 - e^{-\beta},$$

and we assume infected nodes always recover in exactly 1 timestep. Therefore the sole parameter is $\lambda \in (0, 1)$, and we call the critical value $\lambda_c$. The process on a finite graph with $n_0$ nodes is defined by the state vector $X_n \in \{0, 1\}^{n_0}$, $n \geq 0$. As before, a value of 1 corresponds to an infected node and 0 corresponds to susceptible. If a node $i$ is susceptible and has $M$ infected neighbors at time $n$,



(a) $G_1$ (b) $\lambda = 0.12 < (\rho(G_1))^{-1}$ (c) $\lambda = 0.18 > (\rho(G_1))^{-1}$

Figure 5.1: (a) A graph $G_1$ on 50 nodes, grown via preferential attachment with average out-degree 3, spectral radius $\rho(G_1) = 7.8$ and $\frac{1}{\rho(G_1)} = 0.128$. (b) Lifetime of SIS model on $G_1$ in the quick extinction regime. Histogram over 500 runs with a single, randomly selected, initially infected node and $\lambda = 0.12 < (\rho(G_1))^{-1}$. (c) Lifetime of SIS model on $G_1$ in the slow extinction regime. Histogram over 500 runs with a single, randomly selected, initially infected node and $\lambda = 0.18 > (\rho(G_1))^{-1}$. Simulations are truncated at 100 timesteps.

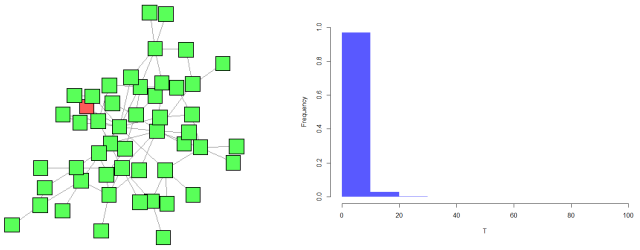
then $i$ is infected at time $n + 1$ with probability $1 - (1 - \lambda)^M$. We denote the lifetime of the process by

$$\tau = \inf\{n \geq 0 : X_n = 0\},$$

the first time step in which all nodes are susceptible.

The lifetime of a discrete-time spreading process under the SIS model is a random variable with values in $[1, \infty)$ and dependent on the transmission probability $\lambda$. In general, the larger $\lambda$, the longer the process is likely to survive. See, for example, Figure 5.1, to see the empirical distribution of lifetime for two values of $\lambda$: one in the regime of quick extinction regime, the other slow extinction.

Figure 5.1 intimates that the lifetime distribution may possibly be heavy-tailed in the slow extinction regime, at least for certain types of networks, a fact which is alluded to be never stated in, e.g., [69, 33]. It is important to remember however, that there is always a spread of lifetime values. In particular, even for very large values of $\lambda$, there is always a chance of extinction in the first few timesteps.




(a) $G_2$ (b) $\lambda = 0.18 < (\rho(G_2))^{-1}$

Figure 5.2: (a) A graph $G_2$ on 50 nodes, grown via preferential attachment with average out-degree 2, spectral radius $\rho(G_2) = 4.7$ and $\frac{1}{\rho(G_2)} = 0.213$. (b) Lifetime of SIS model on $G_2$ in the slow extinction regime. Histogram over 500 runs with a single, randomly selected, initially infected node and $\lambda = 0.18 < (\rho(G_2))^{-1}$. Simulations truncated at 100 timesteps.

It is also important to emphasize the role of graph structure in the lifetime of spreading processes. In general, graphs that are more dense or "more connected" in any sense, may likely have longer lifetimes. See Figure 5.2 for a secondary example on a graph with lower edge density.

Comparing $G_1$ and $G_2$ in Figures 5.1 and 5.2, we see that lifetime does not just depend on transmission rate $\lambda$, but also on the graph in question. In a sense, we can use the spectral radius $\rho$ as a measure of the "spreadability" of a graph, where the larger the spectral radius, the easier the spreading and longer the expected lifetime.

### 5.1.2 Related Work

The original work in this chapter examines the lifetime of a spreading process, but often we are interested in a different quantity: the prevalence $r$ of an infection. The prevalence at time $t$ is the proportion of nodes that are infected at time $t$, that is, $r(t) = \frac{1}{n_0} \sum_{i=1}^{n_0} X_i(t)$ [68]. It is know that on infinite graphs, the SIS

model approaches a steady-state solution as $t \to \infty$ with constant prevalence. In general, we aim to solve a stochastic spreading model by computing the distribution of $r(t)$ for all $t \geq 0$.

The SIS (or any similar) model can be exactly solved on any specific graph using a Markov chain [71, 86]. A Markov chain is constructed with state space $\{S, I\}^N$, where $N$ is the number of nodes in the network. States gives the infection status of each node in the graph, and transition probabilities must be derived according to the adjacency matrix. The primary limitations of this approach are in computational expense, as the state space is of size $2^N$. Some work has been done to reduce the dimensionality to linear in $N$, using a lumped system that only considers the number of infected nodes [80].

An additional drawback is that this process must be repeated for each graph instance, as opposed to covering a class of graphs, e.g., Erdős Rényi, or preferential attachment type graphs.

Solving exact stochastic models for epidemic spreading is notoriously challenging. Newman develops an exact solution to the SIR model is via generating functions, in relation to percolation theory [63]. However, for the SIS model, exact solutions are lacking. Often, an approximation based on the mean-field method is used to formulate a system of deterministic differential equations [4]. Although there is a mathematical basis for this type of approximation [52, 53], recent work makes little attempt to justify its use mathematically [69, 11, 9]. Rather, the approximations are empirically compared to simulation results, often only through a visual plot.

Further, using a finite graph simulation to estimate steady-state behavior on

an infinite graph is inherently questionable. In response to this issue, we initially began studying growing graphs as a potential method for estimating the steady-state prevalence of the asymptotic graph. Recent work has shown that for complete graphs on an increasing number of nodes, the expected infected proportion converges to the solution of the mean-field ODE model [71, 80, 5]. These results are encouraging, but more work needs to be done to prove the general case.

New work in community detection has also prompted the question of epidemic models on graphs with communities. Recent work is largely empirical in nature, and we discuss these works in more detail in Section 5.2. See [68, 48, 88] for more thorough reviews of classical and recent approaches to epidemic models.

## 5.2 SIS Model on Networks with Communities

This section aims to contribute to the ongoing investigation of how community structure affect spreading processes. As stated in Section 1.1, many real-world networks are organized into communities of strongly-interconnected nodes, and therefore it is important to understand the implications of this structure on network processes.

Recent work is largely empirical in nature, with some mean-field approaches. Studies such as [89, 82, 57] utilize graph models that allow for variation in community structure, cluster coefficients, or other measures, and compare SIS or SIR simulations on these graphs in order to conjecture some trend or correlation. Others (e.g., [58, 44]) compare graphs with community structure to

graphs that maintain many of the same characteristics (i.e., degree distribution), but do not have community structure.

The general consensus is that community structure inhibits epidemic spread, because, while an infection may spread rapidly within and permeate an infected community, it is unlikely to spread to other communities, because of the relative sparsity of connecting edges. Thus the overall infected population may be small compared to a graph without community structure, and the lifetime of the infection shorter [44, 89]. Correspondingly, effective intervention policies target nodes that bridge communities, rather than simply nodes with high degree or centrality [77].

However, not all studies came to this conclusion. In fact, some reported the opposite effect, that community structure actually assists epidemic spreading [58], and others simply found no inhibition effect [19, 82]. Stegehuis et. al found that the presence of communities either helped or inhibited epidemic spread, depending on other structural properties [83].

The inconsistency of results suggests that some other factor is at play here, and that there isn't one clear effect of community structure in general.

Of course, the results depend on the epidemic model used, as the recurrent behavior of SIS and SI is quite different from the transitivity of SIR. However there are contradictory results even within the same model (e.g., SIS in [58, 82], SIR in [89, 77, 83, 57], and SI [19, 44]).

Also to be noted is that each study uses a different graphical model, from Erdős-Rényi types to Preferential Attachment type models, and each have their own method for creating communities within these models. More worrisome,

some studies compare community graphs to control (community-free) graphs that are quite different in terms of degree distribution [58, 57]. One must be careful not to attribute cause to the presence of communities when there are other possible explanations, e.g., higher variance in node degree.

Further, there is no universally accepted definition of community, nor a trustworthy measure of the strength of community in a graph (see Section 1.1.1 for a discussion on the limitations of modularity as a community measure). Therefore I would not be surprised if the presence of communities or not is not even a determining factor of any particular spreading effect. I worry that there is not much insight to be gained when we are just reporting correlations between community strength and ease of spreading for a particular combination of epidemic model, graph model, and notion of community.

Stegehuis et. al does the most to understand the role of community in epidemic spreading, and what types of structures are important in determining spreading behavior. In my opinion, their most cogent result is that, if communities are relatively dense, then the internal structure of the communities is not a strong influence on the spreading process [83].

My aim is to make a small contribution to the work of identifying when and how communities are an important factor to consider in analyzing epidemic models. I focus entirely on the SIS model, and report results that are valid for any graph, without reliance on any particular definition or measure of community. This is done by throwing out any reference to communities, and considering, instead, a partition of a graph. I will refer to *groups* of nodes defined by this partition, which may be connected via edges in a way that qualitatively resembles communities, but this will not be assumed.

The idea is to consider how the SIS model behaves on the subgraphs defined by the graph partition, and compare this behavior to that on the graph as a whole. Qualitatively, our simulation studies corroborate observations stated in many previous studies: if an infection is introduced in a group such that the transmission rate $\beta$ dominates the epidemic threshold for that group, then the infection may persist in that group for some time, periodically entering any neighboring groups at a rate that is dependent on the connections between neighboring groups. (See [70] for a detailed simulation study and mean-field analysis of this effect on a variant of the stochastic block model with two communities.) Further, we analytically relate the epidemic threshold of an entire graph to the epidemic thresholds of its subgraphs, under any graph partition. Our result shows that the epidemic threshold of the entire graph is closely related to the epidemic threshold of the group with largest spectral radius, and specifies the dependency on edges connecting to other groups.

This result assists in making sense of previous results. For example, Li et. al consider a model where each communities has a different edge density. They state that as "heterogeneity among communities" increases, then epidemic threshold decreases. However, in their model, an increase in "heterogeneity" actually corresponds to an increase in the density of their most dense community [57]. Thus, they are increasing the spectral radius of this community, which, according to our Theorem 18, will decrease the epidemic threshold.

### 5.2.1 Main Result

We will denote $\rho(A)$ as the spectral radius, that is, the largest eigenvalue, of an adjacency matrix $A$ associated with a graph $G$. An important result due to

Ganesh et. al states that on the SIS model with transmission rate $\beta$, if $\beta \leq \frac{1}{\rho(A)}$, then the epidemic dies out quickly, with expected lifetime on the order of $\log n$, where $n$ is the number of nodes in $G$. Thus the spectral radius $\rho(A)$ characterizes, to some extent, the epidemic threshold and the lifetime of the infection.

Suppose that an undirected graph $G$ on $n$ nodes is partitioned into $k \geq 1$ groups, defining subgraphs $G_1, \ldots, G_k$, where $G_i$ has $n_i$ nodes, $i = 1, \ldots, k$. Write the adjacency matrix $A$ of $G$ as

$$A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1k} \\ A_{21} & A_{22} & & \\ \vdots & & \ddots & \vdots \\ A_{k1} & & \cdots & A_{kk} \end{pmatrix}, \tag{5.1}$$

where $A_{ii}$ is the $n_i \times n_i$ adjacency matrix of each subgraph $G_i$, and $A_{ij}$ is the $n_i \times n_j$ matrix representing edges connecting group $i$ and group $j$, $i, j = 1, \ldots, k$. Write any eigenvector $v$ of $A$ as $v = (v_1, \ldots, v_k)$, where $v_i = (v_{i,1}, \ldots, v_{i,n_i})$ is a vector of length $n_i$.

This formulation allows us to understand the spectral radius of $A$ in terms of the spectral radii associated with the subgraphs $G_1, \ldots, G_k$ and the edges between groups.

**Theorem 18.** *Suppose $\rho(A_{ii}) \geq \rho(A_{jj})$ for some $i$ and all $j = 1, \ldots, k$. Then*

$$\rho(A_{ii}) \leq \rho(A) \leq \rho(A_{ii}) + \delta, \tag{5.2}$$

*where, letting $v$ be the normalized eigenvector associated with the eigenvalue $\rho(G)$ of $A$,*

$$\delta = 2 \sum_{1 \leq j < \ell \leq k} \sum_{s=1}^{n_j} \sum_{t=1}^{n_\ell} v_{j,s} v_{\ell,t} A_{j\ell}(s,t). \tag{5.3}$$

Note that the sum in (5.3) is taken over all edges that connect two distinct groups, and that $v_{j,s} \geq 0$ for any $j, s$ by the Perron-Frobenius theorem [60]. Letting $C$ be the total number of connecting edges in $G$, we see that a trivial upper bound for (5.3) is $\delta \leq C$.

Thus, joining multiple groups of nodes (communities) together increases the spectral radius, but only up to a point. The increase in spectral radius is bounded by a quantity that depends upon the number of connecting edges. Since spectral radius is correlated with expected lifetime in the SIS model, this result suggests that ease of spreading depends mostly on the dominant group of nodes (community) and the connecting edges, rather than other measures of community structure.

The proof uses the following fact from spectral graph theory, which is a restatement of Lemma 4 in [87].

**Lemma 12.** *Suppose a set of edges $M$ is removed from a graph $G$, to form the graph $G_M$. Let $A$ be the adjacency matrix of $G$ and $v$ the normalized eigenvector associated with $\rho(A)$. Let $A_M$ be the adjacency matrix of $G_M$ and $u$ the normalized eigenvector associated with $\rho(A_M)$. Then*

$$2 \sum_{(i,j)\in M} u_i u_j \leq \rho(A) - \rho(A_M) \leq 2 \sum_{(i,j)\in M} v_i v_j.$$

*Proof.* See [87]. □

*Proof of Theorem 18.* Recall (5.1) and consider the block diagonal matrix

$$B = \begin{pmatrix} A_{11} & 0 & \cdots & 0 \\ 0 & A_{22} & & \\ \vdots & & \ddots & \vdots \\ 0 & & \cdots & A_{kk} \end{pmatrix}.$$

The eigenvalues of $B$ are simply the set of eigenvalues associated with $A_{11}, \ldots, A_{kk}$. Therefore the largest eigenvalue of $B$ is

$$\rho(B) = \max_{j=1,\ldots,k} \rho(A_{jj}) = \rho(A_{ii}).$$

Let $u_i$ be the normalized eigenvector associated with the eigenvalue $\rho(A_{ii})$ of $A_{ii}$. Then the normalized eigenvector associated with $\rho(B)$ is $u = (0, \ldots, 0, u_i, 0, \ldots, 0)$.

Let $M$ be the set of all between-community edges of $G$. Then $B$ is the adjacency matrix of the graph resulting from removing the edges in $M$ from $G$. Note that every edge in $M$ is represented by a 1 in one of the off-diagonal matrices $A_{j\ell}$, $j \neq \ell$. Thus Lemma 12 states

$$\rho(A) \leq \rho(B) + \sum_{1 \leq j < \ell \leq k} \sum_{s=1}^{n_j} \sum_{t=1}^{n_\ell} v_{j,m} v_{\ell,n} A_{j\ell}(s,t)$$

and

$$\rho(A) \geq \rho(B) + \sum_{1 \leq j < \ell \leq k} \sum_{s=1}^{n_j} \sum_{t=1}^{n_\ell} u_{j,m} u_{\ell,n} A_{j\ell}(s,t). \tag{5.4}$$

Since $u_{j,m} = 0$ for all $j \neq i$, then $u_{j,m} u_{\ell,n} = 0$ for each term of the sum. Therefore (5.4) becomes

$$\rho(A) \geq \rho(B) = \rho(A_{ii}).$$

□

The relevance of Theorem 18 to the SIS model is seen when we consider the epidemic threshold $\beta_c(A) = \frac{1}{\rho(A)}$.

**Corollary 2.** *In the context of Theorem 18,*

$$\beta_c(A_{ii}) - \delta\left(\beta_c(A_{ii})\right)^2 \leq \beta_c(A) \leq \beta_c(A_{ii}). \tag{5.5}$$

Thus, by joining groups of nodes (communities) together, the epidemic threshold decreases, but it decreases at most by a second-order term of the threshold $\beta_c(A_{ii})$ of the strongest community.

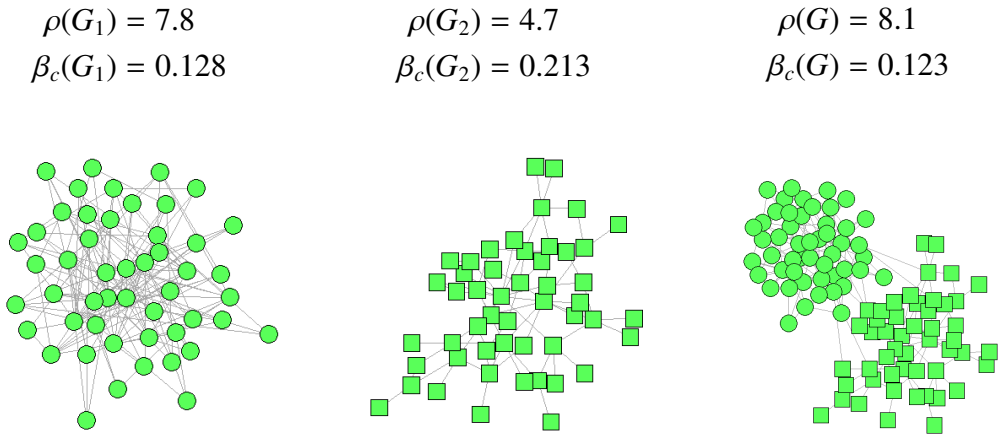
The lower bound is nontrivial when $\beta_c(A_{ii}) < 1/\delta$, i.e., when $\rho(A_{ii}) > \delta$. In practice, I have observed this condition to hold for graphs generated by the CA-PAM model, which is intended to produce graphs with community structure. I hope to show that, in general, $\rho(A_{ii}) > \delta$ for graphs with communities, or, alternately, perhaps this comparison could provide a measure to the strength of community structure in a graph.

### 5.2.2 Numerical Example

Here we give a numerical example to illustrate Theorem 18 and understand the epidemic model on a small graph with two communities.

Recall $G_1$ and $G_2$ from Figures 5.1 and 5.2. We formed $G$ by connecting them with 7 edges, as in Figure 5.3c. Spectral radii and epidemic thresholds were calculated for all three graphs, as seen in Figure 5.3. In this case, $G_1$ has larger spectral radius than $G_2$, and the combined graph $G$ has only slightly larger spectral radius than $G_1$. Accordingly, the epidemic threshold $\beta_c(G)$ is slightly lower than $\beta_c(G_1)$, only differing in the third decimal.

We ran discrete-time SIS simulations with transmission rate $\lambda = 0.18$ on $G_1$, $G_2$, and $G$. Recalling Figure 5.1c, we know that it is possible for the infection to



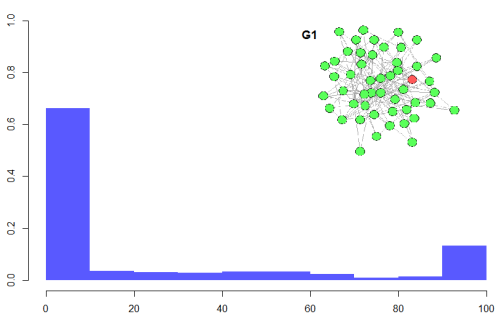

(a) $G_1$ (b) $G_2$ (c) $G = G_1 \cup G_2 \cup C$

Figure 5.3: Comparisons of spectral radius and epidemic threshold for $G_1$, $G_2$, and $G$.

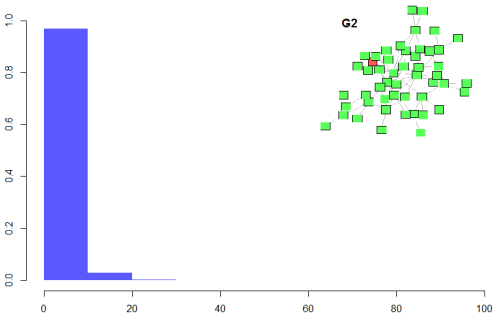
persist for a relatively long time (over 100 timesteps) on $G_1$, which is consistent with the fact that $\lambda = 0.18$ is above its epidemic threshold. On the other hand, $\lambda$ is below the epidemic threshold for $G_2$, and we see that simulations always die out quickly in Figure 5.2b.

What happens, then, when we run the same simulation on $G$? In this case, $\lambda = 0.18$ is still above the epidemic threshold, so we expect to see some long lifetimes, but as we will see, epidemic spread on $G$ is still limited by the structure of $G_2$.

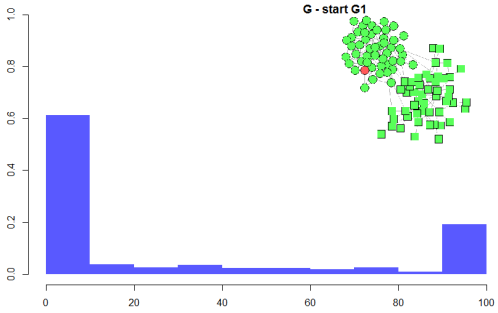
See Figure 5.4 for empirical distributions of the lifetime our simulations on the various graphs, and with various initial conditions. In general, we observe that the behavior of the infection depends mostly on the structure of the community in which the infected nodes belong, and that the connecting edges contribute to extending the lifetime, but not enough to fundamentally change the lifetime distribution.



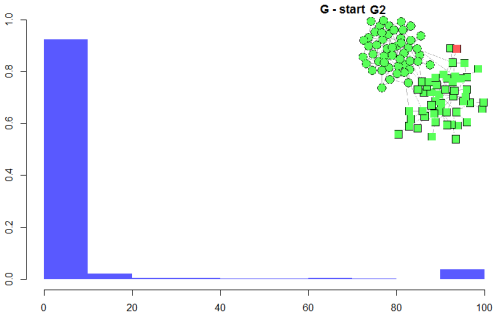

(a) on $G_1$ only



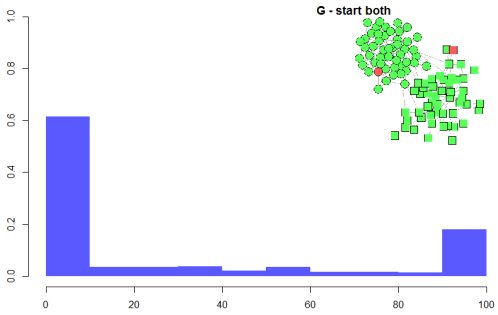

(b) on $G_2$ only



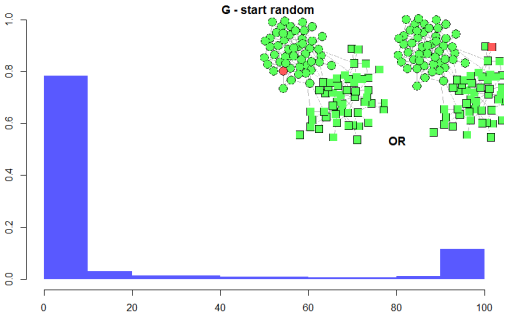

(c) on $G$, but initial infected node in $G_1$.




(d) on $G$, but initial infected node in $G_2$.




(e) on $G$, with one initial infected node in $G_1$ and $G_2$.




(f) on $G$, with initial infected node randomly selected.

Figure 5.4: Histograms of extinction times on $G_1, G_2$, and $G$ with $\lambda = 0.18$

Consider Figure 5.4c, which reports a histogram of the lifetime of infections in $G$ when the infection originates in a randomly selected node from $G_1$. In this case, the initial behavior is as if $G_1$ were the entire graph. It is possible that the infection persists for a long time in $G_1$, in which case it periodically sends the infecton across the connecting nodes to $G_2$. Once in $G_2$, however, the infection is limited by the connectivity structure of $G_2$, and is unlikely to survive for long. It

is also possible that infected nodes in $G_2$ return the favor, and reinfect $G_1$ nodes, but this seems to have little impact on the overall lifetime. In Figure 5.4c, we see that the empirical lifetime distribution is almost identical to the distribution on $G_1$ alone (Figure 5.4a), except for a slight increase in lifetime across the board.

When the initially infected node belongs to $G_2$ (Figure 5.4d), the infection is likely to die out quickly. Sometimes the infection manages to cross over into $G_1$, in which case we have the same behavior as if the infection started in $G_1$. This is represented in Figure 5.4d but the fact that the empirical lifetime distribution looks very much like that on $G_2$ alone (Figure 5.4b), except for a few cases when it manages to survive past the 30 timestep mark.

In Figure 5.4e, the infection is initialized on one node from each of the two communities, and we see a very similar empirical distribution as when it was only introduced in $G_1$. In Figure 5.4f, we initiate at a random node from either community, and the empirical distribution looks like a mixture of the $G_1$-start and the $G_2$-start cases.

### 5.2.3 Conclusion

While this simulation study is certainly limited by size, number of communities, and graph model, I hope this gives insight to behavior of spreading processes on graphs with communities. We have seen in these simulations that spreading behavior is strongly characterized by the connectivity structure of the individual communities, and that the connecting edges certainly increase the average lifetime, but only up to a point.

This observation is supported by Banerjee, et. al in their finding on the limit

of the influence of external agents on an epidemic model [7]. Outside intervention (or intervention from a connecting community) necessarily prolongs lifetime and assists spreading, but when it is bounded in some sense, due to a limiting number of connecting edges, for example, then the effect of this assistance is also limited.

Although this is only one contrived example, Figure 5.3 illustrates the result of Theorem 18 that the spectral radius (and epidemic threshold) of a graph may be very close to the corresponding quantities in the dominant community. In this case, the fact of community structure, or the strength of the communities, may be less indicative of spreading behavior than the structure of the individual communities. Also the lifetime distribution can change substantially depending on where the infection originates, and this can be explained by considering the epidemic threshold of the community or group of origin.

One next step of this project is to formalize and prove a conjecture that the gap $\delta$ as in (5.3) disappears as $C/n \to 0$, where $n$ is the number of nodes in the graph, and $C$ is the number of edges connecting distinct groups in the graph partition. Another step is to examine the spectral radius of a graph resulting from randomly rewiring the edges of a community graph. This direction would be much more involved than the present work, as the spectral radius is random in this case, and we may possibly need to specify a model for the community graph.

A future work may wish to extend this project to the SIR model. I would conjecture quite different behavior in this case, as the lifetime has a hard upper bound and there is no recurrent behavior. I believe, however, that the same ground-up approach will be effective for studying this case, as one can consider

how the process spreads throughout a single commuity, and the likelihood of escaping that community, depending on the connecting edges.

## 5.3 SIS Model on Growing Networks

Recent efforts have been made on time-varying or temporal networks, in which nodes and edges exist for only a subset of the entire timeline [26, 62, 85]. We consider an SIS model on a specific class of temporal preferential attachment networks: those which are strictly increasing in time.

The theoretical result due to Ganesh et. al on the epidemic threshold $1/\rho$ are limited to static graphs, which do not change over time [33]. Our class of growing networks is thus interesting because, while all graphs in the sequence are finite, the limiting graph is infinite, and thus this class exists between the boundary of finite and infinite graphs. We will show that there is an intriguing behavioral difference of SIS-type processes on growing versus static networks.

It is important to note the results of [71, 80, 5], which prove that, on a complete graph that grows over time, the dynamics of a SIS model converges to the mean-field approximation. As yet, these results have not been extended to general graphs to the best of our knowledge, but they do lend credibility to our conjecture that infinite survival is possible on growing graphs.

We use simulations to compare the lifetime distributions of SIS processes on growing graphs versus finite static graphs generated via preferential attachment. Our computational results show that the distribution of an infection's lifetime has a heavier tail in the growing case compared to the static case. Con-

trary to the static case, the empirical lifetime distribution may be classified as long-tailed. Further, for large enough values of the spreading rate $\lambda$, we give evidence that infinite survival may be possible on the growing graph. We also argue that, unlike static graphs, the epidemic threshold for long survival does not depend on initial graph size.

### 5.3.1 Time-varying networks

The vast majority of epidemic research considers spreading processes on static graphs: graphs that remain fixed as the virus propagates, and have no interaction with the spreading process [68]. More generally, one may consider graphs that change over time, sometimes called temporal or time-varying graphs [42]. These graphs may not be strictly increasing or decreasing over time. For example, an email communication network may contain different connections during work hours than during evening hours. These types of networks, where edges disappear and reappear periodically, are discussed in [68], and a mean-field method has been developed for more general classes of time-varying networks in [36].

One interesting study considers the interaction of SIR spreading on a preferential attachment network where nodes and edges arrive over time, but nodes are removed (along with their corresponding edges) after being infected [26]. Interestingly, the remaining nodes were seen to have a degree distribution that decreased from a power law to exponential as infection rate increased.

In this work we focus our attention on the class of preferential attachment graphs that are increasing with time and do not interact with the spreading pro-

cess. Nodes and edges arrive independently according to the usual preferential attachment rules, and the infection may access new edges and nodes as soon as they are available. While this may be seem a simpler model than the aforementioned examples, this model allows us to more directly understand the effect of network growth on a SIS-type spreading process.

We consider an SIS model on a specific class of temporal networks: those which are strictly increasing in time according to a graph sequence $G = (G_0, G_1, G_2 \ldots)$, where $G_0$ has $n_0$ nodes and $m_0$ edges, and $G_{n+1}$ is formed by adding exactly one node and $m \geq 1$ edges to $G_n$.

Our purpose in studying growing networks is two-fold. First is the direct application to real networks in the era of ever-increasing information. The second is the potential use of growing network simulations to estimate the steady-state prevalence of the SIS model on infinite graphs. We come to no conclusions on this second idea, as we have yet to show that a growing network even has a steady-state solution. Rather, our results give evidence that infinite survival is possible, which is necessary for a steady-state solution to exist.

In the discrete-time SIS model, nodes exist in one of two possible states, susceptible and infected. A susceptible node may become infected at the next time step if it is adjacent to an infected node. The probability that the infected node transmits the infection to the susceptible node is given by a parameter $\lambda$. The higher the value of $\lambda$, the longer the infection persists in the network. In the discrete-time case, we will denote the epidemic threshold as $\lambda_c$, where for $\lambda < \lambda_c$ the infection dies out quickly, and for $\lambda > \lambda_c$ the infection may survive for a "long" time [33]. The notion of "long" depends on the cardinality of the node set, as the SIS model on an infinite graph may survive for an infinite amount of

time, while a finite graph necessarily dies out in finite time.

The question of this work is then, what does "long" survival mean on the growing network? Is it possible for the SIS model to survive infinitely on a growing network, and if so, is there a critical threshold for infinite survival? We aim to examine the intriguing behavior of the lifetime of the SIS model on the growing graph, which exists somewhere between finite and infinite static graphs.

We use simulations to compare the lifetime distributions of SIS processes on growing graphs versus finite static graphs. Our computational results show that the distribution of an infection's lifetime has a heavier tail in the growing case compared to the static case. Contrary to the static case, the empirical lifetime distribution can be classified as long-tailed. Further, for large enough values of the spreading rate $\lambda$, we give evidence that infinite survival may be possible on the growing graph. We also argue that, unlike static graphs, the epidemic threshold for long survival does not depend on initial graph size.

### 5.3.2 Model

Consider a graph sequence $G = (G_0, G_1, G_2, \ldots)$ where $G_0$ has $n_0$ nodes and $G_n$ is a subgraph of $G_{n+1}$ for all $n \geq 0$. We call this a growing or increasing graph sequence, and define a discrete-time SIS model on the graph sequence as follows:

- Initialize $X_0 \in \{0, 1\}^{n_0}$.
- If $X_n(i) = 1$ at time $n \geq 0$, then set $X_{n+1}(i) = 0$.
- If $X_n(j) = 0$ or $j$ is born at step $n + 1$, let $M$ be the number of neighbors $i$ of

$j$ in $G_{n+1}$ such that $X_n(i) = 1$,

$$M := \sum_i A_{n+1}(i, j)\mathbb{1}(X_n(i) = 1),$$

where $A_{n+1}$ is the adjacency matrix of $G_{n+1}$. Set

$$X_{n+1}(j) = \begin{cases} 1 & \text{with probability } 1 - (1 - \lambda)^M \\ 0 & \text{with probability } (1 - \lambda)^M. \end{cases}$$

- The extinction time, or lifetime, of the process is

$$\inf\{n \geq 0 : X_n(i) = 0 \text{ for all } i = 1, 2, \ldots, n_0 + n\}.$$

We consider a growing graph sequence generated by the preferential attachment model. Letting $PA_0$ be the graph with a single node and zero edges, we generate recursively. Given $PA_n$, we form $PA_{n+1}$ by adding a new node and forming $m$ edges between this node and existing nodes. The existing nodes are selected with probability proportional to their degree.

In our simulations, unless stated otherwise, we use $n_0 = 1000$, $m = 1$ and let

$$G_0 = PA_{n_0-1}$$
$$G_n = PA_{n_0-1+n}, \qquad n = 1, 2, 3, \ldots.$$

Note that $G_n$ has $n_0 + n$ nodes and $m(n_0 + n - 1)$ edges. Simulations are run until extinction or up to $n = 5000$ timesteps (final graph size is $n_0 + 5000 = 6000$), whichever comes first. At least 1000 iid observations of termination time were recorded for values of $\lambda = 0.10, 0.11, \ldots, 0.21$.

It is clear that growth must only prolong, not shorten, the lifetime. Extinction time in the static graph $G_0$ is stochastically dominated by extinction time in $G =$

$(G_0, G_1, \ldots)$. Therefore the static graph gives a lower bound on extinction time and probability of infinite survival, and an upper bound on the critical value for long survival.

We conjecture, however, that shifting an SIS process from a static graph to a growing graph sequence causes more than just a constant or linear increase in these values, rather, it causes fundamental changes to the lifetime distribution and to the notion of critical value. Intuitively, the longer the process survives, the larger the graph, and hence the easier the spreading and the more likely it is to continue to survive. This contributes to a heavier-tailed lifetime distribution and the possibility of infinite survival. Further, it diminishes the notion of a critical spreading rate, because critical rates tend to decrease as graph size increases.

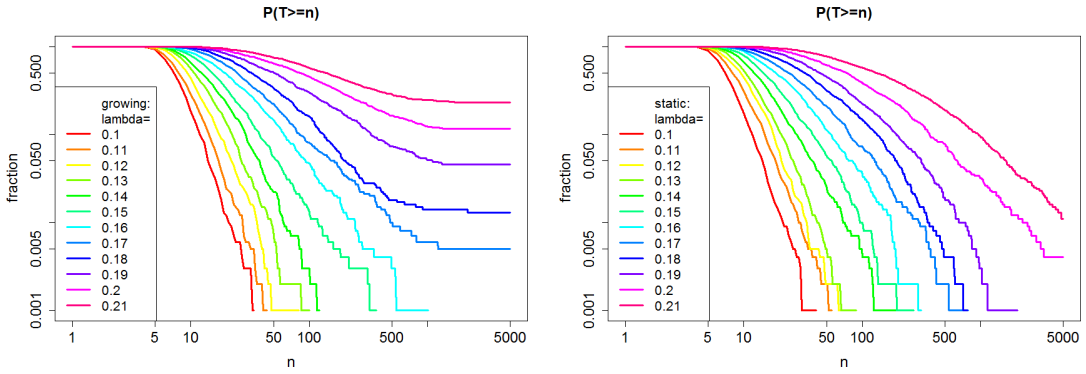
Our computational results lead us to believe that on $G$:

- infinite survival is possible;
- lifetime distribution has a heavier tail compared to that on a static graph;
- critical value for slow extinction does not depend on initial graph size.

In the growing scheme, we must clarify what we mean by critical value. Do we mean a threshold for infinite lifetime, or a threshold for long survival? Does this value depend on the size of the initial graph? The different notions of critical value are addressed in the last section.

### 5.3.3 Evidence for infinite survival

First, we summarize our data in Fig. 5.5, which shows the complementary cdfs of the lifetime of an SIS process on a growing graph (top) and on a static graph (bottom). We notice that the cdf in the growing case may follow a power law, for at least some values of $\lambda$, and suggests a possibly infinite tail for $\lambda \geq 0.17$.



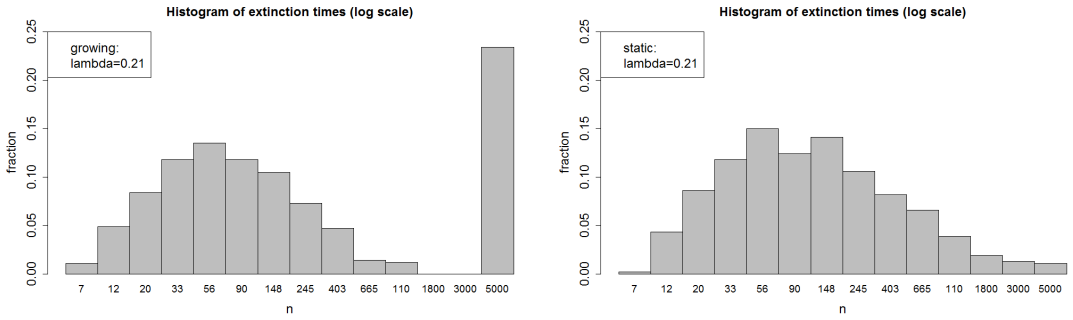

(a) (b)

Figure 5.5: Complementary cumulative distribution functions for the lifetimes of SIS processes on (a) a growing graph and (b) a static graph. For boths graph types, the spreading processes have been run for a range of spreading probabilities $\lambda = 0.10, 0.11, \ldots, 0.21$.

#### 5.3.3.1 Mass at $\infty$

One surprising result was seen on the growing graph for large values of $\lambda \geq 0.17$. Once the process had survived for a certain amount of time, it seemed to always survive until termination. Empirically, on simulations terminated at 5000 timesteps, we never observed extinction times in the interval $(2000, 5000)$. Letting $T$ be the lifetime of our simulated process, there was always a value $\eta << 5000$ such that

$$P(T \geq 5000 \mid T \geq \eta) = 1.$$




(a) (b)

Figure 5.6: Histograms of SIS log-lifetime with $\lambda = 0.21$ on (a) a growing sequence of graphs and (b) a static graph.

This suggests that if a process survives for a certain number of time steps, then the graph has now grown large enough that future extinction is highly unlikely.

When termination was increased to 10,000 time steps, the same phenomenon occurred, with no extinctions in the interval $(2000, 10000)$. We believe this is evidence that such long-surviving processes will survive indefinitely.

On the other hand, on static graphs, we saw extinctions at any point up until termination. Specifically, there were extinctions at around 4000, as seen in the histograms in Fig. 5.6. This shows that in the static case, conditioning on long survival does not guaranteed continued survival.

To test the hypothesis of potentially infinite survival in the growing case, we ran one long simulation of 50,000 time steps. The infection prevalence for this process was recorded every ten time steps and plotted in Fig. 5.7, where the prevalence at time $n$ is defined to be the fraction of infected nodes at time $n$.

We see that prevalence became quite stable over time, which suggests that prevalence may continue to stabilize around this rate as $n \to \infty$. This indicates a possible steady-state where, once reached, the infection will never become

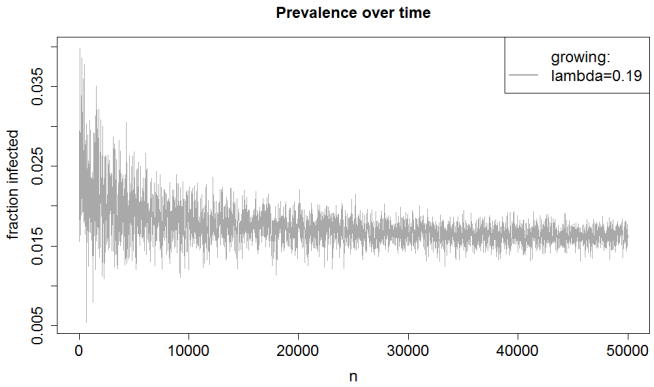



Figure 5.7: Fraction of infected nodes at time $n$ for $n = 10k$, $k = 1, 2, \ldots, 500$ on a growing graph sequence with $\lambda = 0.19$.

extinct.

#### 5.3.3.2 Conditional survival probability

Letting $\tau$ denote the lifetime of an SIS process, we consider

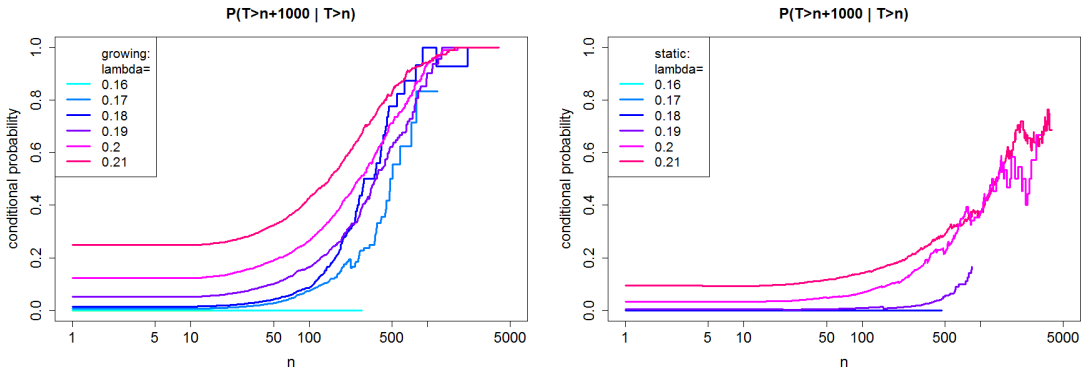
$$P(\tau > n + b \mid \tau > n), \tag{5.6}$$

the probability that the infection will survive for an additional $b$ timesteps, given that the infection is still present at time $n$. Our main observation for spreading processes on a growing graph is that the conditional survival probability (5.6) tends to increase as $n \to \infty$. That is, the longer the infection has survived, the more likely it is to continue to survive.

This quantity is directly related to long-tailed distributions. We say that a random variable $X$ has a **long-tailed** distribution if for all $t > 0$

$$\lim_{x \to \infty} P(X > x + t \mid X > x) = 1.$$

The above statement will be true for any distribution with a mass at infinity, in addition to some non-infinite heavy-tailed distributions.




(a) (b)

Figure 5.8: Empirical $P(\tau > n + 1000 | \tau > n)$ for SIS processes on (a) growing graph sequences and (b) a static graphs.

In our simulations, we fix a value of $\lambda$ and we generate 1000 independent graph sequences $G^{(i)} = (G_0^{(i)}, G_1^{(i)}, \ldots)$, $i = 1, \ldots, 1000$. For each graph sequence $G^{(i)}$, we run the discrete-time SIS model with transmission probability $\lambda$ and record the extinction/termination time $\tau_i$, $i = 1, \ldots, 1000$. In Figure 5.8a, we estimate (5.6) for $b = 1000$ by plotting

$$\frac{\sum_{i=1}^{1000} \mathbb{1}(\tau_i > n + 1000)}{\sum_{i=1}^{1000} \mathbb{1}(\tau_i > n)} \tag{5.7}$$

for $n = 1, \ldots, 1000$. This process is repeated for several values of $\lambda$.

In Figure 5.8b, we plot a similar estimate for the SIS model on static graphs. In this case, for each value of $\lambda$, we generate 1000 independent preferential attachment graphs $PA_{n_0-1}^{(i)}$, $i = 1, \ldots, 1000$. We run the discrete-time SIS model on each of the *static* graphs $PA_{n_0-1}^{(i)}$, record the extinction/termination time $\tau_i'$, and plot the fraction resulting from putting $\tau_i'$ in place of $\tau_i$ in (5.7).

In Figure 5.8, in both the static and growing cases, we see an increase in the conditional probability as $n$ increases. However, the conditional probability increases faster with $n$ in the growing case than it does in the static case. Further, as $n \to 5000$, the conditional probability appears to converge to 1, whereas in the

static case, the limit seems to be less than 1. In the growing case, this is evidence of a long-tailed distribution and consistent with mass at infinity.

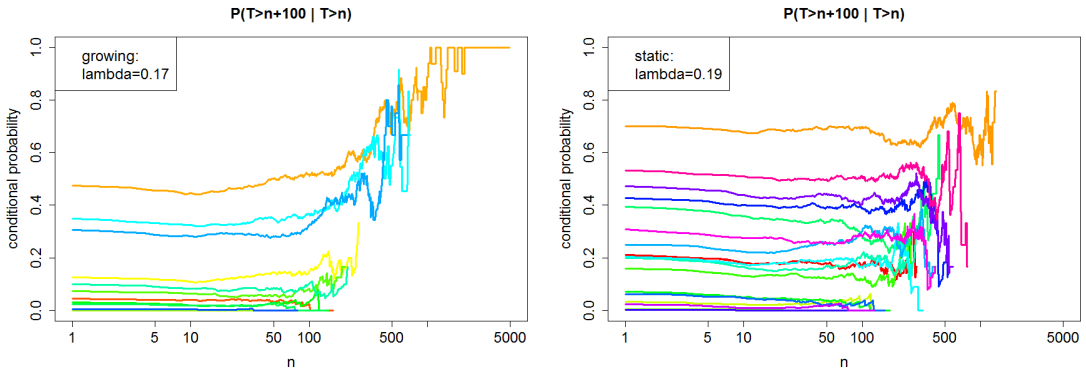
We must consider, however, an alternate explanation to the increasing slopes in both parts of Figure 5.8. Each point in the plots is calculated from $\tau_i$ ($\tau_i'$) values based off of 1000 independent graph sequences (static graphs). It is possible that there are some especially communicable graphs on which an infection is likely survive longer than it would on another graph. If an infection has survived for at least $n$ time steps, then it is more likely to be acting on a "communicable" graph, and therefore more likely to survive for a very long time. Therefore, $\tau_i'$ ($\tau_i$) may be positively correlated with some "communicability" property of graph $i$. This correlation may cause a spurious increase with $n$ in the calculation of (5.7).

In short, the issue is that we only run a single SIS simulation on each graph (graph sequence). To understand whether this truly causes a noticeable effect, we ran another set of simulations in which 1000 SIS simulations were run on a single graph. That is, we fixed $\lambda = 0.17$, generated $G = (G_0, G_1, \ldots)$, and ran 1000 discrete-time SIS simulations this single instance of $G$. Letting $\tau_1, \ldots, \tau_{1000}$ denote the extinction/termination times, we then plotted

$$\frac{\sum_{i=1}^{1000} \mathbb{1}(\tau_i > n + 100)}{\sum_{i=1}^{1000} \mathbb{1}(\tau_i > n)} \tag{5.8}$$

for $n = 1, \ldots, 5000$ in Figure 5.9a. This process was repeated 12 times in total, corresponding to the 12 distinctly-colored curves in Figure 5.9a.

In the static case, we fixed $\lambda = 0.19$, generated $PA_{n_0-1}$, and recorded the extinction/termination times $\tau_1', \ldots, \tau_{1000}'$ of 1000 discrete-time SIS simulations on $PA_{n_0-1}$. We then plotted the equivalent of (5.8) for $n = 1, \ldots, 5000$ in Figure 5.9b.




(a) (b)

Figure 5.9: Empirical $P(\tau > n + 100|\tau > n)$ for SIS processes with (top) $\lambda = 0.17$ on a growing graph and (bottom) $\lambda = 0.19$ on a static graph. Each curve corresponds to a distinct graph or graph sequence.

It should be noted that in Figure 5.9, we are only looking $b = 100$ steps into the future, as opposed to $b = 1000$ in Figure 5.8. This is because, for each of the 12 instances of $G$ and $PA_{n_0-1}$ in Figure 5.9, it was very rare for the infection to survive longer than 1000 time steps, and so taking $b = 1000$ does not yield informative results. In comparison, in the data for Figure 5.8 there were over 1000 instances each of $G$ and $PA_{n_0-1}$, and on a few of the SIS simulations, the lifetime was longer than 1000 time steps. This supports our claim of "communicability": that there are some graphs and graph sequences that are especially advantageous for the infection.

In Figure 5.9b, we see that, for a given static graph, the conditional survival probability $P(\tau' > n + 100 \mid \tau' > n)$ remains more or less constant with $n$. The curves remain flat and perhaps even decrease as $n$ increases. This is quite different from the previous plot, and indicates that the increasing curves of Figure 5.8b may indeed be spurious, caused solely by differences in the simulated graphs.

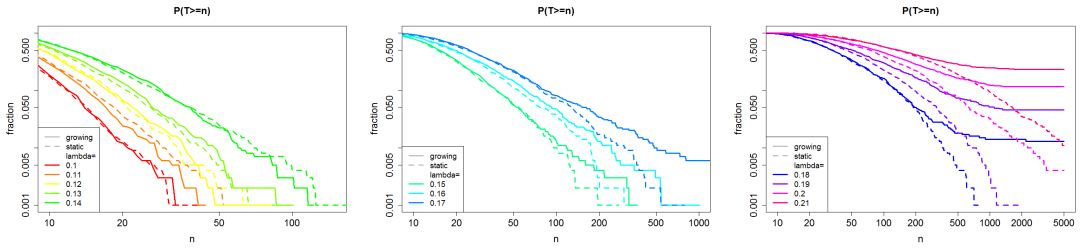
For growing graphs, as in Figure 5.9a, we also see interesting behavior. In many of the simulations, the curve remains flat, while in others, there is an increase as $n$ increases. Therefore we can conclude that for some, but not all, graph sequences, the conditional survival probability $P(\tau > n + 100 \mid \tau > n)$ may increase with $n$. In fact, the sharp increase in Figure 5.8a could result from an averaging effect over the curves in Figure 5.9a. Therefore, in the growing case, our alternate explanation accounts for some of the increase, but not all. We believe that Figure 5.9a still shows evidence of long tails, but that the lifetime distribution may be strongly graph-dependent.

In summary, while the estimates of conditional survival probability (5.6) in Figure 5.8 may not be entirely reliable, the secondary simulations (shown in Figure 5.9) show that there is indeed a difference in conditional survival probability between the growing and static cases. On some growing graphs, we have seen that the longer an infection has survived, the more likely it is to continue to survive. Meanwhile, on static graphs, survival up to a certain time does **not** imply greater probability of continued survival. So we may conclude that for infections on growing graphs, the lifetime distribution has longer tails, as well as heavier tails, compared to infections on static graphs.

### 5.3.4 Lifetime distributions

Recall the empirical ccdfs of lifetime in Fig. 5.5. To compare the distributions in the growing and static cases more closely, we consider three regions of $\lambda$. First, consider Fig. 5.10a, where the tails are plotted for both distributions for $\lambda \leq 0.14$.

For these smallest values of $\lambda$, we see virtually no difference between the




(a) $\lambda$ in the first region. (b) $\lambda$ in the second region. (c) $\lambda$ in the third region.

Figure 5.10: Tails of lifetime distributions in the growing (solid) and static (dashed) cases.

two cases. This is consistent with the explanation that for $\lambda$ small enough, the process doesn't live long enough to gain the advantage of the growing graph. We conjecture that there is some first critical value $\lambda_1$, where if $\lambda < \lambda_1$, there is negligible difference between SIS processes on the growing graph and those on the static graph.

Above this critical value, however, there appears to be significant difference in the two cases. In Fig. 5.10b, for example, the distribution on the growing graph appears to have heavier tails than the distribution on the static graph. In fact, lifetime in the growing case seems to follow a power law.

In Fig. 5.10c, we compare distributions for the largest values of $\lambda$, and clearly see a difference between the two cases. We conjecture a second critical value, $\lambda_2 \geq \lambda_1$, which marks the boundary for distributions of a third type in the growing case. We believe that in the third case, SIS processes on a growing sequence of graphs have possible infinite survival.

### 5.3.5 Epidemic Threshold

Often, we are interested in a critical transmission probability that marks the threshold for "long survival". Whatever the definition of long survival, it is usually the case that the so-called epidemic threshold $\lambda_c$ tends to decrease as size of the graph $n$ increases. For this reason, we now argue that these threshold values have minimal relevance to spreading processes on a growing sequence of graphs, because no matter the transmission rate or the size of the initial graph, the process may theoretically survive for long enough that the graph becomes sufficiently large and the relevant threshold becomes sufficiently small.

Suppose that there is some critical $\lambda_n$ for each $n$, such that if $\lambda > \lambda_n$, then a SIS process on a static graph of size $n$ survives for a long time. Let there be an SIS process with rate $\lambda$ on a sequence of growing graphs $G$ that initializes at $G_0$ with size $n_0$. If $\lambda \geq \lambda_{n_0}$, clearly lifetime will be "long" on the growing graph as well.

On the other hand, suppose $\lambda_{n_0+k} < \lambda < \lambda_{n_0}$ for some $k$ potentially large. Then long survival is possible as long as the virus survives on the growing graph for at least $k$ timesteps. This implies that the critical value $\lambda_c$ for $G$ is actually

$$\lambda_c = \liminf_{n\to\infty} \lambda_n. \tag{5.9}$$

On preferential attachment networks, the epidemic threshold has been stated to vanish with an increasing number of nodes [9]. In this case, $\liminf_{n\to\infty} \lambda_n = 0$, and so the growing preferential attachment graph would theoretically have no epidemic threshold. This means that for any values of $n_0 \geq 1$ and $\lambda \in (0, 1)$, there is a positive probability $p(\lambda; n_0)$ of long survival. Certainly,

$p(\lambda; n_0)$ may be quite small for small values of $\lambda$, since survival to $k$ timesteps may be highly unlikely. In this case, this notion of epidemic threshold may be of little relevance to practical applications.

For infections on static networks, the epidemic threshold has been defined as the transmission rate above which long survival is possible [33]. In the context of growing networks, however, we posit that it may not be expedient to merely consider the possibility of long survival. Depending on the application, there may be a more useful threshold value, such as a transmission rate corresponding to a critical point in the likelihood of long survival.

### 5.3.6 Empirical epidemic threshold

We ran simulations on growing preferential attachment and Erdős-Rényi networks to estimate the epidemic threshold $\lambda_c$. The parameters are $\alpha \in \{0, 1\}, m \in \{1, 2, 3\}$, and $\xi \in \{0, 1\}$.

Graphs are generated from $PA_0$ via preferential attachment with power $\alpha = 0$ or 1, so that the probability of forming an edge with a node of degree $d$ is proportional $d^\alpha$. One new node and $m = 1, 2$ or 3 new edges are added at each time step.

Continue in this manner until the graph has 1000 nodes. After that, we introduce the infection on 50% of the nodes, and then if $\xi = 1$, we continue to grow the graph in the same manner, so that we alternate between: adding a new node with $m$ edges; and progressing one step in the SIS model. If $\xi = 0$, we continue to progress in the SIS model, but the graph does not grow.

| | |
|---|---|
| $\rho$ | spectral radius: largest eigenvalue of adjacency matrix |
| $\eta$ | isoperimetric constant |
| $\lambda < \frac{1}{\rho}$ | sufficient condition for quick extinction |
| $\lambda > \frac{1}{\eta}$ | sufficient condition for long survival |
| $1/\hat{\eta}$ | my estimate (lower bound) for $\frac{1}{\eta}$. |

Table 5.1: Summary of notation and theoretical conditions

What values of $\lambda$ do we use for each graph?

Our results are summarized in Table 5.2. We report the empirical critical rate $\hat{\lambda}_c$ as the minimal value of $\lambda$ where we witness survival up to 5000 time steps. Compare to theoretical lower and upper bounds for static graphs $1/\rho$, $1/\eta$ as in [33]. We also compare maximal degree, diameter, and cluster coefficient. Averages over at least 1000 simulations.

The notation ER stands for the static Erdős-Rényi graph (selected uniformly over all graphs with 1000 nodes and 1000 edges).

$$\eta = \min_{S \subset \{1,2,\ldots,n\}, |S| \leq n/2} \frac{\text{cut}(S, S^c)}{|S|}, \qquad \hat{\eta} = \frac{\text{cut}(\hat{S}, \hat{S}^c)}{|\hat{S}|} \geq \eta, \qquad 1/\hat{\eta} \leq 1/\eta.$$

$\hat{S}$ = smallest of 2 communities detected via fast-greedy algorithm

All graphs initialized at 1000 nodes. After virus introduced, $\xi$ new nodes added each time step with $m$ new edges each, connected to $w$ with probability proportional to $w^\alpha$.

Note: $\xi = 0$ implies static graph, $\alpha = 1$ corresponds to linear preferential attachment, $\alpha = 0$ is dynamic erdos-renyi, ER is static erdos-renyi.

E.g., PA(1,1) corresponds to power $\alpha = 1$, $\xi = 1$, $m = 1$, avg degree = 2.

| ER or Power $\alpha$ | Growth rate $\xi$ | Avg deg $2m$ | Crit rate $\hat{\lambda}_c$ | Est crit 1 $1/\rho$ | Est crit 2 $1/\hat{\eta}$ | Max deg $d_{\max}$ | Diam. $D$ | Local CC $c_\ell$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 2 | 0.17 | 0.12 | 355 | 67.0 | 17.3 | 0 |
| 1 | 1 | 4 | 0.09 | 0.067 | 0.998 | 170 | 6.8 | 0.087 |
| 1 | 1 | 6 | 0.06 | 0.046 | 0.519 | 256 | 5.0 | 0.16 |
| 0 | 1 | 2 | 0.31 | 0.23 | 367 | 13.3 | 21.7 | 0 |
| 0 | 1 | 4 | 0.17 | 0.14 | 1.94 | 21.6 | 8.05 | 0.0071 |
| 0 | 1 | 6 | 0.12 | 0.097 | 0.498 | 28.8 | 6.2 | 0.010 |
| 1 | 0 | 2 | 0.21 | | | | | |
| 0 | 0 | 2 | 0.33 | | | | | |
| ER | 0 | 2 | 0.35 | 0.29 | 4.4 | 7.9 | 22.2 | 0.0020 |

Table 5.2: Empirical epidemic threshold compared to theoretical values, for three types of random graphs.

9 total graphs, each run on about 8 values of $\lambda$. For each graph and $\lambda$, at least 1000 simulations.

Considering Table 5.2, $1/\rho$ looks to be a reasonable lower bound The estimated upper bound $1/\hat{\eta}$ is very loose, and $1/\eta$ can only be more loose. Recall that these bounds, $1/\rho$, etc., are for static graphs. We conjectured that the growing thresholds should actually be lower than the static thresholds, however there is no clear way to compare, as we don't know true static thresholds, and if the probability of infinite survival is vanishingly small for vanishing lambda, then we are unable to run enough simulations to show it.

### 5.3.7 Infinite survival on the growing star graph

The star graph on $n$ leaves is the graph consisting of one central node of degree $n$ and $n$ nodes of degree 1 (called leaves) , where each leaf's sole edge connects to the central node. Here we give a quick proof that infinite survival of the

SIS model is possible on a growing star graph, given that the number of leaves increases sufficiently fast over time.

Let $S = (S_0, S_1, \ldots)$ be the growing sequence of graphs where $S_n$ is the star graph with $mn + n_0$ leaves, $n \geq 0$. Suppose the center node is infected at time 0 and then the SIS model progresses with transmission rate $\lambda$. Let $C_n$ be the random indicator that the center node is infected at time step $2n$, and let

$$T = \inf\{n \geq 0 : C_n = 0\}$$

denote the lifetime of the process.

**Lemma 13.** *Suppose $\lambda < 1$ and $n_0, m$ are large enough to satisfy $2(1-\lambda)^{n_0} + \frac{(1-\lambda)^{\frac{1}{4}\lambda n_0}}{1-(1-\lambda)^{\frac{1}{4}\lambda m}} < 1$. Then*

$$P(T < \infty) < 1.$$

*Proof.* Define the binomial random variables

$$B_n \sim Bin(mn + n_0, \lambda), \qquad n = 0, 1, 2, \ldots.$$

Suppose $C_n = 1$ for some $n \geq 0$, that is, the center node is infected at time $2n$. Then we say that the number of infected leaves at time $2n + 1$ is $B_n$. If there is a positive number of infected leaves at time $2n + 1$, then the center node will be infected at time $2n + 2$ with probability $P(C_{n+1} = 1) = 1 - (1 - \lambda)^{B_n}$.

Let $\mathcal{B}$ be the event that $B_0, B_1 \geq n_0$ and $B_n \geq \lambda(mn + n_0) - \sqrt{\lambda(mn + n_0)\log n)} > 0$ for all $n \geq 2$. We will show that $P(\mathcal{B}) > 0$.

By Hoeffding's inequality,

$$P(B_n < \lambda(mn + n_0) - \sqrt{\lambda(mn + n_0)\log n}) \leq \exp\left(-\frac{2(\lambda(mn + n_0)\log n)}{\lambda(mn + n_0)}\right) = \frac{1}{n^2},$$

and therefore

$$\begin{aligned}
&P(B_n < \lambda(mn+n_0) - \sqrt{\lambda(mn+n_0)\log n} \text{ for any } n \geq 2) \\
&\leq \sum_{n=2}^{\infty} P(B_n < \lambda(mn+n_0) - \sqrt{\lambda(mn+n_0)\log n}) \\
&\leq \sum_{n=2}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6} - 1.
\end{aligned}$$

Since $P(B_n \geq \lambda(mn+n_0) - \sqrt{\lambda(mn+n_0)\log n) \text{ for all } n \geq 2) = 1 - P(B_n < \lambda(mn+n_0) - \sqrt{\lambda(mn+n_0)\log n} \text{ for any } n \geq 2)}$, we have

$$\begin{aligned}
P(\mathcal{B}) &= P(B_0 = n_0)P(B_1 \geq n_0)P(B_n \geq \lambda(mn+n_0) - \sqrt{\lambda(mn+n_0)\log n}) \text{ for all } n \geq 2) \\
&\geq (\lambda^{n_0})^2 \left(2 - \frac{\pi^2}{6}\right) > 0.
\end{aligned}$$

Thus it is sufficient to show that $P(T < \infty \mid \mathcal{B}) < 1$. We have

$$P(C_n = 0 \mid B_{n-1}, C_1 = \cdots = C_{n-1} = 1) = (1-\lambda)^{B_{n-1}}.$$

Note:

$$x - \sqrt{x\log(x)} > \left(1 - \frac{1}{\sqrt{2}}\right)x > \frac{1}{4}x$$

for all $x \geq 1$. Letting $x = \lambda(mn+n_0)$, then $\lambda(mn+n_0) - \sqrt{\lambda(mn+n_0)\log n} > x - \sqrt{x\log x}$, so

$$\begin{aligned}
P(T < \infty \mid \mathcal{B}) &\leq \sum_{n=1}^{\infty} P(C_n = 0 \mid \mathcal{B}) \\
&= P(C_1 = 0 \mid B_0 = n_0) + P(C_2 = 0 \mid B_1 \geq n_0) + \sum_{n=3}^{\infty} P(C_n = 0 \| B_{n-1}) \\
&\leq 2(1-\lambda)^{n_0} + \sum_{n=2}^{\infty} (1-\lambda)^{\lambda(mn+n_0) - \sqrt{\lambda(mn+n_0)\log n}} \\
&\leq 2(1-\lambda)^{n_0}) + \sum_{n=2}^{\infty} (1-\lambda)^{\frac{1}{4}\lambda(mn+n_0)} \\
&\leq 2(1-\lambda)^{n_0} + (1-\lambda)^{\frac{1}{4}\lambda n_0} \frac{1}{1-(1-\lambda)^{\frac{1}{4}\lambda m}}.
\end{aligned}$$

□

As should be clear in the above proof, the condition $(1-\lambda)^{n_0} + \frac{(1-\lambda)^{\frac{1}{4}\lambda n_0}}{1-(1-\lambda)^{\frac{1}{4}\lambda m}} < 1$ is sufficient, but not necessary. I believe the result requires some condition guaranteeing large $n_0$ and/or $m$. Next steps would be to find a minimal condition and find the exact probability of infinite survival.

I also hope to prove the same result for the continuous-time SIS model, and then use this to prove that infinite survival is possible on growing preferential attachment graphs.

### 5.3.8 Conclusion

We have given evidence that network growth causes a fundamental shift in the distribution of the lifetime of an infection under the SIS model. The addition of new nodes and edges over time may allow the infection to spread further and survive longer than the static case, resulting in a lifetime distribution with heavier tails compared to the static case.

For very small values of the transmission rate parameter $\lambda$, the effect of growth on the lifetime distribution is minimal, as the infection is likely to die out before it has a chance to take advantage of the new connections. However, for larger values of $\lambda$, possibly above some threshold $\lambda_1$, the infection may survive for a longer time. If it survives long enough so that the graph has grown sufficiently big, then spreading may become even easier over time. Thus, the longer an infection has survived, the less likely it is to die out in the immediate future. This contributes to heavy-tailed behavior in the lifetime distribution, and possibly long-tailed behavior. For $\lambda$ above a secondary threshold, $\lambda_2$, this effect may be strong enough that the infection persists indefinitely.

We also discuss that the epidemic threshold $1/\rho$, defined for the SIS model on static graphs, may be of minimal utility in the growing case, because the threshold is graph-dependent and decreases as the graph increases.

Future steps would be to prove mathematically the conditions for infinite survival on growing preferential attachment graphs and to study the behavior of infection prevalence as $n \to \infty$. If infinite survival is possible, then the SIS model on the growing graph may converge to a steady-state, and the steady-state infection prevalence on the growing graph may be related to the steady-state infection prevalence on the infinite graph, which can be difficult to estimate via simulation. If there is a relation between steady-state prevalence on growing and infinite graphs, then the growing graph may be a useful tool to estimate spreading behavior on infinite graphs.

APPENDIX A

# COUPLING $G$ AND $\hat{G}$

## A.1 Coupling Overview

In this section we couple the graph sequences $G = (G(t), t \geq 0)$ and $\hat{G} = (\hat{G}(t), t \geq 0)$. We do this to prove Lemma 5.

Recall

$$t^* = \sup\{t \in \mathbb{Z}^+ : q - \frac{1+\alpha}{2\alpha}\sqrt{\frac{\log t}{t}} \leq 0\}, \tag{A.1}$$

$$\eta = \min\left\{\frac{2\delta}{t*},\ q - \frac{1+\alpha}{2\alpha}\sqrt{\frac{\log(t^*+1)}{t^*+1}} + 2\delta\right\}, \tag{A.2}$$

and

$$C_X = 3q + \frac{q}{\eta}\frac{1+\alpha}{2\alpha} + \frac{\delta q + \left(\delta + \frac{3}{2}\right)(1-\alpha)|q_{with} - q_{bet}|}{2\sqrt{\alpha}}. \tag{A.3}$$

The definitions for graph sequences $G$ and $\hat{G}$, in the case of two communities, are restated below for reference.

Set $G(0)$ to be the graph with 2 nodes and zero edges with one node in each community. At time $t$, we grow the graph from $G(t)$ to $G(t+1)$ through the following mechanism:

- with probability $\alpha$, a new node is added to $G(t)$ and its community membership is chosen uniformly from $\{1, 2\}$. We assign $V_t$ to be this new node.
- with probability $1 - \alpha$, we assign $V_t$ to be a random node from $G(t)$.

- $V_t$ sends an edge invitation to $W_t$, chosen with probability

$$P(W_t = w) = \frac{D_t^w + \delta}{m(t) + \delta n(t) + 2\delta},$$

  where $D_t^w$ is the in-degree of node $w$ at time $t$, $n(t) + 2$ is the number of nodes in $G(t)$ and $m(t)$ is the total number of edges in $G(t)$.

- with probability $q_{V_t,W_t}$, $W_t$ accepts the invitation and the directed edge $(V_t, W_t)$ is formed.

Set $\hat{G}(0) = G(0)$. At any time $t \geq 1$, $\hat{G}(t)$ has exactly $\lfloor \alpha t \rfloor + 2$ nodes and $\lfloor qt \rfloor$ edges, and evolves into $\hat{G}(t+1)$ as follows:

- Set $i = 1$ if $\lfloor \alpha(t+1) \rfloor$ is odd, and $i = 2$ if $\lfloor \alpha(t+1) \rfloor$ is even
- If $\lfloor \alpha(t+1) \rfloor > \lfloor \alpha(t) \rfloor$, a new node $\hat{V}_t$ is added to the graph and assigned membership to community $i$. Otherwise, we assign $\hat{V}_t$ to a node randomly selected from community $i$.
- If $\lfloor q(t+1) \rfloor > \lfloor q(t) \rfloor$
  - Assign a value to $J \in \{1, 2\}$ according to

$$P(J = j) = \begin{cases} \frac{q_{with}}{2q} & j = i \\ \frac{q_{bet}}{2q} & j \neq i \end{cases}.$$

  - Given $J$, assign $\hat{W}_t$ according to

$$P(\hat{W}_t = w) = \frac{\hat{D}_t^w + \delta}{\sum_{u \in C_J} (\hat{D}_t^u + \delta)},$$

    and form the directed edge $(\hat{V}_t, \hat{W}_t)$.

## A.2 Coupling Scheme

### A.2.1 Setup and Notation

Define a *community i node* to be a node belonging to community $i$. Nodes are named according to community and order of arrival. E.g., in $G(t)$, there are nodes $0, 1, \ldots, n_1(t)$ belonging to community 1 and nodes $0, 1, \ldots, n_2(t)$ belonging to community 2, for a total of $n(t) = n_1(t) + n_2(t)$ non-initial nodes. Graph $\hat{G}(t)$ has community 1 nodes $0, 1, \ldots, \lfloor \alpha t/2 \rfloor$.

Define a *community i edge* to be an edge with terminal node belonging to community $i$. Let $\tau_k$ be the arrival time of the $k$th community 1 edge in $G$, and let $\hat{\tau}_k$ be the arrival time of the $k$th community 1 edge in $\hat{G}$. Immediately before the arrival of the $k$th community 1 edge in $G$, node $v$ will have in-degree

$$D^v(\tau_k - 1), \qquad v = 0, 1, \ldots, n_1(\tau_k - 1).$$

Immediately before the arrival of the $k$th community 1 edge in $\hat{G}$, node $u$ will have in-degree

$$\hat{D}^u(\hat{\tau}_k - 1), \qquad u = 0, 1, \ldots, \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor.$$

When referring to a node that does not yet exist in a graph, we say that it has in-degree 0. Suppose, for example, that $n_1(t) < u \leq \lfloor \alpha t/2 \rfloor$. Then $u$ exists in $\hat{G}(t)$ but does not exist in $G(t)$. It has some degree $\hat{D}^u(t)$ in $\hat{G}(t)$, and if we must refer to the degree of the corresponding node in $\hat{G}(t)$, we will say $D^u(t) = 0$.

In order to compare the degree distributions in $G$ and $\hat{G}$ at index $k$, we label community 1 nodes as "good" or "bad", depending on whether they have

matching degree in the two graphs. A node $v$ belonging to community 1 is said to be "good" immediately preceding index $k$ if the following two conditions are met

1. Node $v$ exists in both $G(\tau_k - 1)$ and $\hat{G}(\hat{\tau}_k - 1)$, i.e., $v \leq n_1(\tau_k - 1)$ and $v \leq \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor$.
2. Node $v$ has identical in-degree between the two graphs, i.e., $D^v(\tau_k - 1) = \hat{D}^v(\hat{\tau}_k - 1)$.

Thus we define the set of good nodes at index $k$ to be

$$\text{Good}(k) = \left\{ v \in \{0, 1, \ldots, n_1(\tau_k - 1) \wedge \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor\} : D^v(\tau_k - 1) = \hat{D}^v(\hat{\tau}_k - 1) \right\}.$$

When the coupling fails to match terminal nodes at index $k$, then the nodes in question will no longer have identical in-degree between the two graphs, and are thus relegated to the bad set $B(k)$. The bad set $B(k)$ is explicitly defined within the coupling scheme.

There is a third label, disjoint from the labels "good" and "bad". Since the good set at index $k$ only contains nodes that exist in both $G(\tau_k - 1)$ and $\hat{G}(\hat{\tau}_k - 1)$, we define the excess set $E(k)$ to be the set of community 1 nodes $v$ that satisfy the following two conditions:

1. node $v$ exists in one graph, but not the other, immediately preceding the arrival of the $k$th edge, i.e.,

$$v \in \{n_1(\tau_k - 1) \wedge \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor, \ldots, n_1(\tau_k - 1) \vee \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor\}.$$

2. node $v$ has not yet been relegated to the bad set.

Thus,

$$E(k) = \{n_1(\tau_k - 1) \wedge \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor, \ldots, n_1(\tau_k - 1) \vee \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor\} \setminus B(k-1),$$

and any community 1 node that exists in at least one of the graphs at index $k$ belongs to exactly one of the sets Good($k$), $B(k)$, or $E(k)$.

### A.2.2 Detailed Procedure

We will refer to four subroutines which are detailed in section A.2.3: `Node arrivals`; `Community 1 edge arrivals`; `Select` $U_k$ `given` $\hat{U}_k$; and `Select` $\hat{U}_k$ `given` $U_k$.

In the outline below, note that the order of steps 5 and 6 are not interchangeable.

Set $\hat{G}(0) = G(0)$ to be the graph with a single node belonging to community 1, a single node belonging to community 2, and zero edges. Set $B(0) = \emptyset$.

Given $\mathcal{F}_{t-1}$ and $\hat{\mathcal{F}}_{t-1}$, do the following at time $t \geq 1$:

1. Do subroutine `Node arrivals` to decide values of $V_t$ and $\hat{V}_t$.
2. Do subroutine `Community 1 edge arrivals` to decide values of $X_t$ and $Y_t$.
3. **Possibly add community 2 edge to $G$:**

   With probability

$$\frac{\sum_{u \in C_2}(D^u(t-1) + \delta)}{m(t-1) + \delta n(t-1) + 2\delta} \cdot \begin{cases} q_{bet} & \text{if } C(V_t) = 1 \\ q_{with} & \text{if } C(V_t) = 2, \end{cases} \tag{A.4}$$

add the edge $(V_t, W_t)$ to $G(t-1)$, where $W_t$ is selected from community 2 according to

$$P(W_t = w) = \frac{D^w(t-1) + \delta}{\sum_{u \in C_2}(D^u(t-1) + \delta)}, \qquad w = 0, 1, \ldots, n_2(t-1).$$

4. **Possibly add community 2 edge to $\hat{G}$:**

   If $\lfloor qt \rfloor > \lfloor q(t-1) \rfloor$ and $Y_t = 0$, add the edge $(\hat{V}_t, \hat{W}_t)$ to $\hat{G}(t-1)$, where $\hat{W}_t$ is chosen from community 2 with probability

$$P(\hat{W}_t = w) = \frac{\hat{D}^w(t-1) + \delta}{\sum_{u \in C_2}(\hat{D}^u(t-1) + \delta)}, \qquad w = 0, 1, \ldots, \lfloor \alpha(t-1) \rfloor - \lfloor \alpha(t-1)/2 \rfloor.$$

5. **Possibly add community 1 edge to $G$:**

   If $X_t = 1$, set $k = \sum_{s=1}^{t} X_s$ and $\tau_k = t$. Assign $U_k$ to a node from community 1 according to the following.

   (a) The following three statements are equivalent:

   - $\hat{G}(t-1)$ has not yet added its $k$th community 1 edge
   - $\sum_{s=1}^{t-1} Y_s < k$
   - $\tau_k \leq \hat{\tau}_k$

   If the above statements are true, then choose $U_k$ according to

$$P(U_k = w \mid \tau_k = t \leq \hat{\tau}_k) = \left\{ \frac{D^w(t-1)+\delta}{k-1+\delta n_1(t-1)+\delta}, \quad w = 0, 1, \ldots, n_1(t-1). \right.$$

   (b) If case (a) does not hold, then

   - $\hat{G}$ has at least $k$ community 1 edges
   - $\sum_{s=1}^{t-1} Y_s \geq k$
   - $\tau_k > \hat{\tau}_k$

   Let $\hat{U}_k$ denote the terminal node of the $k$th community 1 edge of $\hat{G}$.

   Do subroutine `Select` $U_k$ `given` $\hat{U}_k$ to decide value of $U_k$.

Add edge $(V_t, U_k)$ to $G(t-1)$, and call the resulting graph $G(t)$.

6. **Possibly add community 1 edge to $\hat{G}$:**

   If $Y_t = 1$, set $k = \sum_{s=1}^{t} Y_s$ and $\hat{\tau}_k = t$. Assign $\hat{U}_k$ to a node from community 1 according to the following.

   (a) The following three statements are equivalent:

   - $G(t)$ has not yet added its $k$th community 1 edge
   - $\sum_{s=1}^{t} X_s < k$
   - $\hat{\tau}_k < \tau_k$

   If the above statements are true, then choose $\hat{U}_k$ according to

   $$P(\hat{U}_k = w \mid \hat{\tau}_k = t < \tau_k) = \left\{ = \frac{\hat{D}^w(t-1)+\delta}{k-1+\delta\lfloor\alpha(t-1)/2\rfloor+\delta}, \quad w = 0, 1, \ldots, \lfloor\alpha(t-1)/2\rfloor. \right.$$

   (b) If case (a) does not hold, then

   - $G(t)$ already has at least $k$ community 1 edges
   - $\sum_{s=1}^{t} X_s \geq k$
   - $\hat{\tau}_k \geq \tau_k$

   Let $U_k$ denote the terminal node of the $k$th community 1 edge of $G$.

   Do subroutine `Select` $\hat{U}_k$ `given` $U_k$ to decide value of $\hat{U}_k$.

   Add edge $(\hat{V}_t, \hat{U}_k)$ to $\hat{G}(t-1)$, and call the resulting graph $\hat{G}(t)$.

### A.2.3 Subroutines

`Node arrivals:`

- With probability $\alpha$, add a new node to $G(t-1)$ with community membership chosen uniformly from $\{1,2\}$, and assign $V_t$ to be this new node. With probability $1-\alpha$, assign $V_t$ to be a random node from $G(t-1)$.
- If $\lfloor \alpha t \rfloor > \lfloor \alpha(t-1) \rfloor$, add a new node to $\hat{G}(t-1)$ with community membership 1 if $\lfloor \alpha t \rfloor$ is odd or community membership 2 if $\lfloor \alpha t \rfloor$ is even, and assign $\hat{V}_t$ to be this new node. Otherwise, $\hat{V}_t$ is assigned to a node in $\hat{G}(t-1)$ that is randomly selected from community 1 if $\lfloor \alpha t \rfloor$ is odd or from community 2 if $\lfloor \alpha t \rfloor$ is even.

`Community 1 edge arrivals:`

- Assign a value to $X_t$ according to

$$P(X_t = 1) = 1 - P(X_t = 0) = \frac{\sum_{u \in C_1}(D^u_{t-1} + \delta)}{m(t-1) + \delta n(t-1) + 2\delta} \cdot \begin{cases} q_{with} & \text{if } C(V_t) = 1 \\ q_{bet} & \text{if } C(V_t) = 2. \end{cases} \tag{A.5}$$

  Thus $X_t$ indicates whether a community 1 edge is added to $G$ at time $t$.

- Assign a value to $Y_t$ according to

$$P(Y_t = 1) = 1 - P(Y_t = 0) = \frac{\mathbb{1}_{\lfloor qt \rfloor > \lfloor q(t-1) \rfloor}}{2q} \begin{cases} q_{with} & \text{if } \lfloor \alpha t \rfloor \text{ is odd} \\ q_{bet} & \text{if } \lfloor \alpha t \rfloor \text{ is even.} \end{cases} \tag{A.6}$$

  Thus $Y_t$ indicates whether a community 1 edge is added to $\hat{G}$ at time $t$.

`Select` $U_k$ `given` $\hat{U}_k$

- Fix $k$, $\tau_k = t > \hat{\tau}_k$, and define the following shorthand notation

$$
\begin{aligned}
d_w &= D^w(t-1) + \delta, && w = 0, 1, \ldots, n_1(t-1) \\
\hat{d}_w &= \hat{D}^w(\hat{\tau}_k - 1) + \delta, && w = 0, 1, \ldots, \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \\
d_S &= \sum_{v \in S} (D^v(t-1) + \delta), && S \subset \{0, 1, \ldots, n_1(t-1)\} \\
\hat{d}_S &= \sum_{v \in S} (\hat{D}^v(\hat{\tau}_k - 1) + \delta), && S \subset \{0, 1, \ldots, \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor\} \\
y &= d_{C_1} = k - 1 + \delta n_1(t-1) + \delta \\
\hat{y} &= \hat{d}_{C_1} = k - 1 + \delta \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor + \delta
\end{aligned}
$$

- Given $\hat{U}_k = u$ and all information up to time $t-1$, select $U_k$ according to the following procedure. (The procedure is summarized in the tree diagram on page 156, where the leaves give the conditional probability that $U_k = w$.)

  The procedure is broken into two cases, depending on which graph contains more nodes at index $k$.

**Case 1:** $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1)$.

In this case, the excess set is a subset of $\hat{G}(\hat{\tau}_k - 1)$.

Subdivide this case according to whether $\hat{U}_k$ is currently in the good set Good$(k-1)$ or not. If so, then we try to match $U_k$. If not, then $\hat{U}_k \in B(k-1) \cup E(k-1)$, and we don't worry about matching $U_k$.

**(i)** If $\hat{U}_k = u \in$ Good$(k)$, then $D^u(t-1) = \hat{D}^u(\hat{\tau}_k - 1)$, and we match $U_k$

by assigning it according to

$$P(U_k = w \mid \hat{U}_k = u \in \text{Good}(k),\ \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1),\ \tau_k = t > \hat{\tau}_k) = \begin{cases} 1, & w = u \\ 0, & w \neq u. \end{cases} \tag{A.7}$$

**(ii)** If $\hat{U}_k = u \in B(k-1) \cup E(k)$, then we do not attempt to match $U_k$, and we choose $U_k$ according to

$$P(U_k = w \mid \hat{U}_k = u \in B(k-1) \cup E(k),\ \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1),\ \tau_k = t > \hat{\tau}_k) = \begin{cases} \frac{d_w}{\hat{d}_{B \cup E}} \left( \frac{\hat{y}}{y} - 1 \right), & w \in \text{Good}(k) \\ \frac{d_w}{d_B} \left( 1 - \frac{d_{\text{Good}}}{\hat{d}_{B \cup E}} \left( \frac{\hat{y}}{y} - 1 \right) \right), & w \in B(k-1). \end{cases} \tag{A.8}$$

**Case 2:** $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(t-1)$.

In this case, the excess set is a subset of $G(t-1)$.

Again, we subdivide this case according to whether $\hat{U}_k$ is in the good set $\text{Good}(k)$ or not. If so, then we try to match $U_k$. If not, then $\hat{U}_k \in B(k-1)$, and we don't worry about matching $U_k$.

**(i)** If $\hat{U}_k = u \in \text{Good}(k)$, then $D^u(t-1) = \hat{D}^u(\hat{\tau}_k - 1)$, and we choose $U_k$ according to

$$P(U_k = w \mid \hat{U}_k = u \in \text{Good}(k),\ \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(t-1),\ \tau_k = t > \hat{\tau}_k) = \begin{cases} \frac{\hat{y}}{y}, & w = u \\ \frac{d_w}{d_{B \cup E}} \left( 1 - \frac{\hat{y}}{y} \right), & w \in B(k-1) \cup E(k). \end{cases} \tag{A.9}$$

**(ii)** If $\hat{U}_k = u \in B(k-1)$, we choose $U_k$ according to

$$P(U_k = w \mid \hat{U}_k = u \in B(k-1),\ \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \le n_1(t-1),\ \tau_k = t > \hat{\tau}_k)$$
$$= \frac{d_w}{\hat{d}_B} \frac{\hat{y}}{y} \left(1 - \frac{\hat{d}_{\text{Good}}}{d_{B \cup E}} \left(\frac{y}{\hat{y}} - 1\right)\right), \qquad w \in B(k-1) \cup E(k). \tag{A.10}$$

Lastly, we must update the bad set. It gains new entries if $U_k \neq \hat{U}_k$:

$$B(k) := B(k-1) \cup \{U_k, \hat{U}_k\} \mathbb{1}(U_k \neq \hat{U}_k).$$

Given $\hat{U}_k = u$:

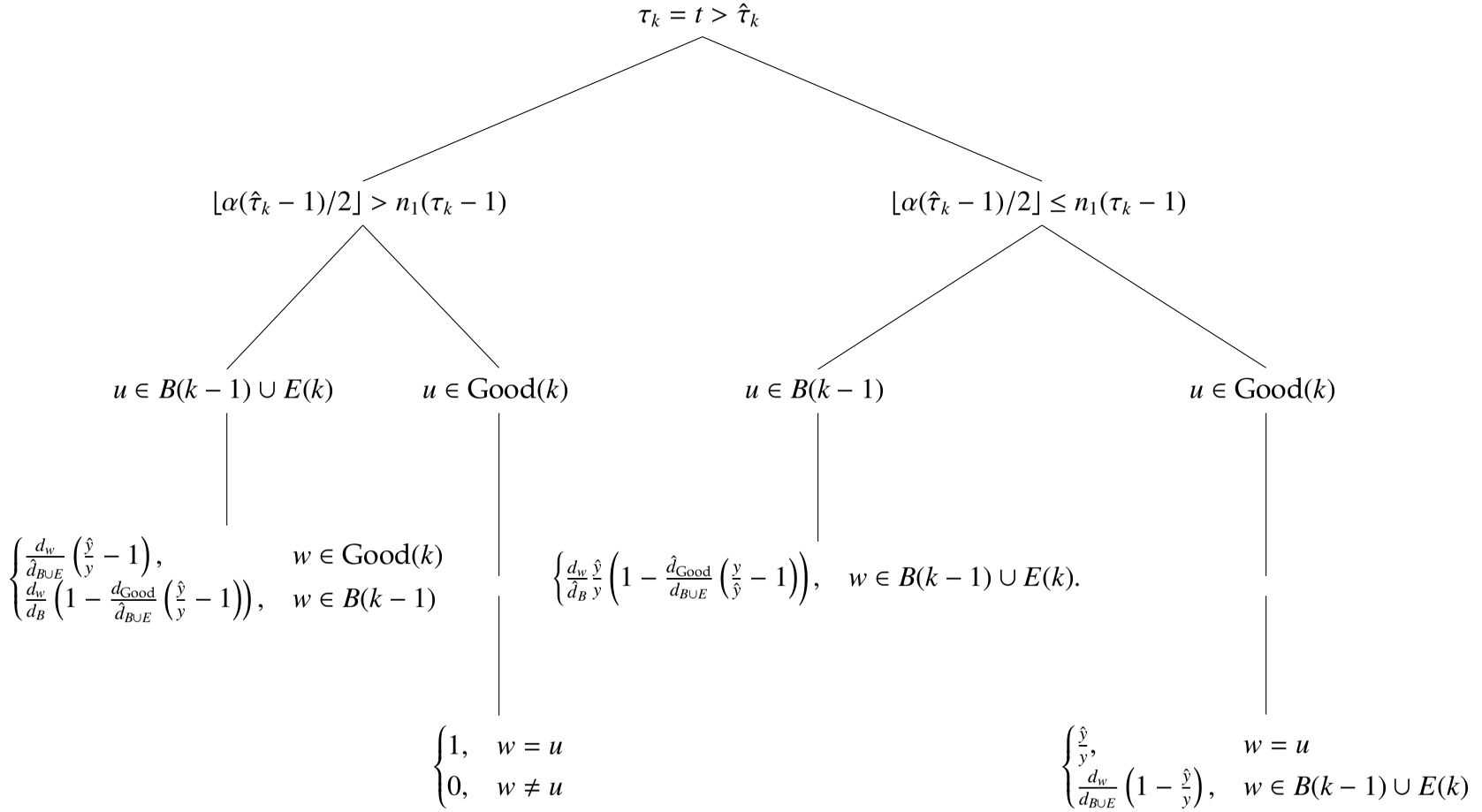



Figure A.1: Diagram depicting coupling procedure under the condition $\tau_k = t > \hat{\tau}_k$. Leaves denote $P(U_k = w)$, conditional upon $\tau_k = t > \hat{\tau}_k$, relative sizes of $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor$ and $n_1(\tau_k - 1)$, and membership of $\hat{U}_k = u$ in Good$(k)$, $E(k)$, or $B(k-1)$.

`Select` $\hat{U}_k$ `given` $U_k$

Analogous to previous, switching roles of $U_k$ and $\hat{U}_k$.

## A.3 Well-definedness

Throughout the proofs we will use the following identities.

**Lemma 14.** *Fix $k \geq 1$ and $\tau_k = t > \hat{\tau}_k$, and refer to the notation in subroutine* `Select` $U_k$ `given` $\hat{U}_k$. *Then*

$$d_{Good} = \hat{d}_{Good} = \begin{cases} y - d_B & \\ y - d_{B \cup E} & \end{cases} = \begin{cases} \hat{y} - \hat{d}_{B \cup E}, & \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1) \\ \hat{y} - \hat{d}_B, & \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(\tau_k - 1) \end{cases}$$

*Proof.* All nodes in Good($k$) have identical degree between $G(\tau_k - 1)$ and $\hat{G}(\hat{\tau}_k - 1)$.

For the graph with a larger number of nodes, the node set is the union of the disjoint sets: Good($k$), $E(k)$, and $B(k-1)$.

For the graph with the smaller number of nodes, the node set is the union of the disjoint sets: Good($k$) and $B(k-1)$. □

### A.3.1 By case in step 5(b)

In step 5(b), within procedure `Select` $U_k$ `given` $\hat{U}_k$, we must prove that the pdf of $U_k$ sums to 1 in each of the four cases. (The same holds for each case in 6(b), within procedure `Select` $\hat{U}_k$ `given` $U_k$.)

**Case 1(i): pdf (A.7)**

$$\sum_w P(U_k = w \mid \hat{U}_k = u \in \text{Good}(k),\ \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1),\ \tau_k = t > \hat{\tau}_k)$$
$$= \sum_w \mathbb{1}(w = u) = 1.$$

**Case 1(ii): pdf (A.8)**

$$\sum_w P(U_k = w \mid \hat{U}_k = u \in B(k-1) \cup E(k),\ \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1),\ \tau_k = t > \hat{\tau}_k)$$
$$= \sum_{w \in \text{Good}(k)} \frac{d_w}{\hat{d}_{B \cup E}} \left( \frac{\hat{y}}{y} - 1 \right) + \sum_{w \in B(k-1)} \frac{d_w}{d_B} \left( 1 - \frac{d_{Good}}{\hat{d}_{B \cup E}} \left( \frac{\hat{y}}{y} - 1 \right) \right)$$
$$= \frac{d_{\text{Good}}}{\hat{d}_{B \cup E}} \left( \frac{\hat{y}}{y} - 1 \right) + \frac{d_B}{d_B} \left( 1 - \frac{d_{Good}}{\hat{d}_{B \cup E}} \left( \frac{\hat{y}}{y} - 1 \right) \right)$$
$$= 1.$$

**Case 2(i): pdf (A.9)**

$$\sum_w P(U_k = w \mid \hat{U}_k = u \in \text{Good}(k),\ \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(t-1),\ \tau_k = t > \hat{\tau}_k)$$
$$= \frac{\hat{y}}{y} + \sum_{w \in B(k-1) \cup E(k)} \frac{d_w}{d_{B \cup E}} \left( 1 - \frac{\hat{y}}{y} \right) = 1.$$

**Case 2(ii): pdf (A.10)** In this case, $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(t-1)$, so by Lemma 14, we

have $\hat{d}_{\text{Good}} = \hat{y} - \hat{d}_B$ and $d_{B\cup E} = y - \hat{y} + \hat{d}_B$.

$$\begin{aligned}
&\sum_w P(U_k = w \mid \hat{U}_k = u \in B(k-1),\ \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \le n_1(t-1),\ \tau_k = t > \hat{\tau}_k) \\
&\qquad = \sum_{w \in B(k-1)\cup E(k)} \frac{d_w}{\hat{d}_B} \frac{\hat{y}}{y} \left(1 - \frac{\hat{d}_{\text{Good}}}{d_{B\cup E}} \left(\frac{y}{\hat{y}} - 1\right)\right) \\
&\qquad = \frac{d_{B\cup E}}{\hat{d}_B} \frac{\hat{y}}{y} - \frac{\hat{d}_{\text{Good}}}{\hat{d}_B} \left(1 - \frac{\hat{y}}{y}\right) \\
&\qquad = \frac{\hat{y}}{y} \left(\frac{y - \hat{y}}{\hat{d}_B} + \frac{\hat{d}_B}{\hat{d}_B}\right) - \frac{\hat{y} - \hat{d}_B}{\hat{d}_B} \left(1 - \frac{\hat{y}}{y}\right) \\
&\qquad = \frac{\hat{y}}{y} \left(\frac{y - \hat{y}}{\hat{d}_B}\right) + \frac{\hat{y}}{y} - \frac{\hat{y}}{\hat{d}_B} \left(\frac{y - \hat{y}}{y}\right) + \left(1 - \frac{\hat{y}}{y}\right) \\
&\qquad = \frac{\hat{y}}{y} + 1 - \frac{\hat{y}}{y} \\
&\qquad = 1.
\end{aligned}$$

### A.3.2 Marginal distributions

**Theorem 19.** *Under the proposed coupling scheme, the following holds true for all $w \in G(t)$:*

$$P(\textit{edge added to } G \textit{ at time } t, \textit{ terminating in } w \mid \mathcal{F}_{t-1}) = \frac{D^w(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta} q_{C(V_t),C(w)}.$$

*Proof.* It suffices to prove the following lemmas, which condition the result over a partition of the graph states and identities of $w$.

| | |
|---|---|
| Lemma 15 | $w \in C_2$ |
| Lemma 16 | $w \in C_1$ and $\tau_k \le \hat{\tau}_k$ |
| Lemma 17 | $w \in C_1$, $\tau_k > \hat{\tau}_k$, and $w \in \text{Good}(k)$ |
| Lemma 18 | $w \in C_1$, $\tau_k > \hat{\tau}_k$, $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1)$, and $w \in B(k-1)$ |
| Lemma 19 | $w \in C_1$, $\tau_k > \hat{\tau}_k$, $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \le n_1(t-1)$, and $w \in B(k-1) \cup E(k)$ |

□

**Lemma 15.** *Conditional upon w belonging to community 2, Theorem 19 holds.*

*Proof.* Fix $w \in \{0, 1, \ldots, n_2(t-1)\}$ in community 2. Then the desired probability is

$$\frac{D^w(t-1)+\delta}{\sum_{u\in C_2}(D^u(t-1)+\delta)} \cdot \frac{\sum_{u\in C_2}(D^u_{t-1}+\delta)}{m(t-1)+\delta n(t-1)+2\delta} \cdot \begin{cases} q_{bet} & \text{if } C(V_t)=1 \\ q_{with} & \text{if } C(V_t)=2 \end{cases}$$
$$= \frac{D^w(t-1)+\delta}{m(t-1)+\delta n(t-1)+2\delta} q_{C(V_t),C(w)},$$

as desired. □

Suppose $w \in \{0, 1, \ldots, n_1(t-1)\}$ in community 1. The probability that an in-edge is added to node $w$ at time $t$, conditional $\mathcal{F}_{t-1}$, is

$$P(\tau_k = t,\ U_k = w \mid \mathcal{F}_{t-1}) = P(X_t = 1 \mid \mathcal{F}_{t-1}) P\left(U_k = w \mid \tau_k = t, \mathcal{F}_{t-1}\right).$$

To calculate $P\left(U_k = w \mid \tau_k = t, \mathcal{F}_{t-1}\right)$, we split into two subcases: either $\tau_k \le \hat{\tau}_k$, or $\tau_k > \hat{\tau}_k$.

**Lemma 16.** *Conditional upon w belonging to community 1 and $\tau_k \le \hat{\tau}_k$, Theorem 19 holds.*

*Proof.* Under the coupling scheme,

$$P(U_k = w \mid \tau_k = t \le \hat{\tau}_k, \mathcal{F}_{t-1}) = \frac{D^w(t-1)+\delta}{k-1+\delta n_1(t-1)+\delta}. \tag{A.11}$$

Then

$$
\begin{aligned}
&P(X_t = 1 \mid \mathcal{F}_{t-1})P\left(U_k = w \mid \tau_k = t \leq \hat{\tau}_k, \mathcal{F}_{t-1}\right) \qquad \text{(A.12)}\\
&= \frac{k - 1 + \delta n_1(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta}\left(\frac{D^w(t-1) + \delta}{k - 1 + \delta n_1(t-1) + \delta}\right) \cdot \begin{cases} q_{with} & \text{if } C(V_t) = 1 \\ q_{bet} & \text{if } C(V_t) = 2 \end{cases}\\
&= \frac{D^w(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta} q_{C(V_t),C(w)}.
\end{aligned}
$$

□

Under subcase $\tau_k > \hat{\tau}_k$, we divide into 3 mutually exclusive subsubcases:

(i) $w \in \text{Good}(k)$

(ii) $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1)$ and $w \in B(k-1)$

(iii) $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(t-1)$ and $w \in B(k-1) \cup E(k)$.

Following the tree on page 156, we now find $P(U_k = w \mid \tau_k = t > \hat{\tau}_k,\ \mathcal{F}_{t-1})$ for each of the subsubcases. We will use the shorthand notation defined in the subprocedure `Select` $U_k$ `given` $\hat{U}_k$. By the same reasoning as in (A.12), it is sufficient to show

$$P(U_k = w \mid \tau_k = t > \hat{\tau}_k,\ \mathcal{F}_{t-1}) = \frac{d_w}{y} \qquad \text{(A.13)}$$

under each of the subsubcases.

**Lemma 17.** *Conditional upon $w$ belonging to community 1, $\tau_k = t > \hat{\tau}_k$, and $w \in$ Good$(k)$, Theorem 19 holds.*

*Proof.* Since $w \in \text{Good}(k)$, we know that $d_w = \hat{d}_w$.

If $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \le n_1(t-1)$, then

$$\begin{aligned}
&P\left(U_k = w \mid w \in \text{Good}(k), \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \le n_1(t-1),\ \tau_k = t > \hat{\tau}_k,\ \mathcal{F}_{t-1}\right) \\
&= \frac{\hat{y}}{y} P\left(\hat{U}_k = w \mid \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \le n_1(t-1),\ \tau_k = t > \hat{\tau}_k\right) \\
&= \frac{\hat{y}}{y} \frac{\hat{d}_w}{\hat{y}} = \frac{d_w}{y}.
\end{aligned}$$

If, on the other hand, $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1)$, then

$$\begin{aligned}
&P\left(U_k = w \mid w \in \text{Good}(k), \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \le n_1(t-1),\ \tau_k = t > \hat{\tau}_k,\ \mathcal{F}_{t-1}\right) \\
&= 1 \cdot P\left(\hat{U}_k = w \mid \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \le n_1(\hat{\tau}_k - 1), \tau_k = t > \hat{\tau}_k\right) \\
&\quad + \frac{d_w}{\hat{d}_{B \cup E}} \left(\frac{\hat{y}}{y} - 1\right) P\left(\hat{U}_k \in B(k-1) \cup E(k) \mid \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \le n_1(\hat{\tau}_k - 1), \tau_k = t > \hat{\tau}_k\right) \\
&= \frac{\hat{d}_w}{\hat{y}} + \frac{d_w}{\hat{d}_{B \cup E}} \left(\frac{\hat{y}}{y} - 1\right) \frac{\hat{d}_{B \cup E}}{\hat{y}} = \frac{\hat{d}_w}{\hat{y}} + \frac{d_w}{y} - \frac{d_w}{\hat{y}} \\
&= \frac{d_w}{y}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&P\left(U_k = w \mid w \in \text{Good}(k),\ \tau_k = t > \hat{\tau}_k,\ \mathcal{F}_{t-1}\right) \\
&= \frac{d_w}{y} P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \le n_1(t-1) \mid \tau_k = t > \hat{\tau}_k,\ \mathcal{F}_{t-1}\right) \\
&\quad + \frac{d_w}{y} P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1) \mid \tau_k = t > \hat{\tau}_k,\ \mathcal{F}_{t-1}\right) \\
&= \frac{d_w}{y},
\end{aligned}$$

and so (A.13) holds in subsubcase (i). □

**Lemma 18.** *Conditional upon $w$ belonging to community 1, $\tau_k = t > \hat{\tau}_k$, $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1)$, and $w \in B(k-1)$, Theorem 19 holds.*

*Proof.* By Lemma 14, we know that $d_{\text{Good}} = y - d_B$ and $\hat{d}_{B\cup E} = \hat{y} - y + d_B$.

$$
\begin{aligned}
&P\left(U_k = w \mid w \in B(k-1),\ \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1),\ \tau_k = t > \hat{\tau}_k,\ \mathcal{F}_{t-1}\right) \\
&= \frac{d_w}{d_B}\left(1 - \frac{d_{\text{Good}}}{\hat{d}_{B\cup E}}\left(\frac{\hat{y}}{y} - 1\right)\right) P\left(\hat{U}_k \in B(k-1)\cup E(k) \mid \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(t-1),\ \tau_k = t > \hat{\tau}_k\right) \\
&= \frac{d_w}{d_B}\left(1 - \frac{d_{\text{Good}}}{\hat{d}_{B\cup E}}\left(\frac{\hat{y}}{y} - 1\right)\right)\frac{\hat{d}_{B\cup E}}{\hat{y}} \\
&= \frac{d_w}{\hat{y}}\frac{\hat{d}_{B\cup E}}{d_B} - \frac{d_w}{\hat{y}}\frac{d_{\text{Good}}}{d_B}\left(\frac{\hat{y}}{y} - 1\right) \\
&= \frac{d_w}{\hat{y}}\left(\frac{\hat{y} - y}{d_B} + \frac{d_B}{d_B}\right) - \frac{d_w}{\hat{y}}\frac{y}{d_B}\left(\frac{\hat{y}}{y} - 1\right) + \frac{d_w}{\hat{y}}\frac{d_B}{d_B}\left(\frac{\hat{y}}{y} - 1\right) \\
&= \frac{d_w}{d_B}\frac{\hat{y} - y}{\hat{y}} + \frac{d_w}{\hat{y}} - \frac{d_w}{d_B} + \frac{d_w}{d_B}\frac{y}{\hat{y}} + \frac{d_w}{y} - \frac{d_w}{\hat{y}} \\
&= \frac{d_w}{d_B} - \frac{d_w}{d_B}\frac{y}{\hat{y}} - \frac{d_w}{d_B} + \frac{d_w}{d_B}\frac{y}{\hat{y}} + \frac{d_w}{y} \\
&= \frac{d_w}{y}.
\end{aligned}
$$

and so (A.13) holds in subsubcase (ii). □

**Lemma 19.** *Conditional upon $w$ belonging to community 1, $\tau_k = t > \hat{\tau}_k$, $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(t-1)$, and $w \in B(k-1)\cup E(k)$, Theorem 19 holds.*

*Proof.* By Lemma 14, we know that $\hat{d}_{\text{Good}} = \hat{y} - \hat{d}_B$

$$
\begin{aligned}
&P\left(U_k = w \mid w \in B(k-1)\cup E(k),\ \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(t-1),\ \tau_k = t > \hat{\tau}_k,\ \mathcal{F}_{t-1}\right) \\
&= \frac{d_w}{d_{B\cup E}}\left(1 - \frac{\hat{y}}{y}\right) P\left(\hat{U}_k \in \text{Good}(k) \mid \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(t-1),\ \tau_k = t > \hat{\tau}_k\right) \\
&\quad + \frac{d_w}{\hat{d}_B}\frac{\hat{y}}{y}\left(1 - \frac{\hat{d}_{\text{Good}}}{d_{B\cup E}}\left(\frac{y}{\hat{y}} - 1\right)\right) P\left(\hat{U}_k \in B(k-1) \mid \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(t-1),\ \tau_k = t > \hat{\tau}_k\right) \\
&= \frac{d_w}{d_{B\cup E}}\left(1 - \frac{\hat{y}}{y}\right)\frac{\hat{d}_{\text{Good}}}{\hat{y}} + \frac{d_w}{\hat{d}_B}\frac{\hat{y}}{y}\left(1 - \frac{\hat{d}_{\text{Good}}}{d_{B\cup E}}\left(\frac{y}{\hat{y}} - 1\right)\right)\frac{\hat{d}_B}{\hat{y}} \\
&= \frac{d_w}{d_{B\cup E}}\left(1 - \frac{\hat{y}}{y}\right)\frac{\hat{d}_{\text{Good}}}{\hat{y}} + \frac{d_w}{y} - \frac{d_w}{y}\frac{\hat{d}_{\text{Good}}}{d_{B\cup E}}\left(\frac{y}{\hat{y}} - 1\right) \\
&= \frac{d_w}{y} + \frac{d_w}{y}\frac{\hat{d}_{\text{Good}}}{d_{B\cup E}}\left(\frac{y - \hat{y}}{\hat{y}}\right) - \frac{d_w}{y}\frac{\hat{d}_{\text{Good}}}{d_{B\cup E}}\left(\frac{y - \hat{y}}{\hat{y}}\right) \\
&= \frac{d_w}{y},
\end{aligned}
$$

and so (A.13) holds in subsubcase (iii). □

**Theorem 20.** *Under the proposed coupling scheme, the following holds true for all* $w \in \hat{G}(t)$*:*

$$P(\text{edge added at time } t, \text{ terminating in } w \,|\, \hat{\mathcal{F}}_{t-1}) = \frac{\hat{D}^w(t-1) + \delta}{\lfloor q(t-1) \rfloor + \delta \lfloor \alpha(t-1) \rfloor + 2\delta} q_{C(\hat{V}_t), C(w)}.$$

*Proof.* Omitted. □

## A.4 Concentration Inequalities

In this section we state and prove various lemmas that will be useful later on, which pertain to $n(t)$, $m_1(t)$, $\hat{m}_1(t)$, $\tau_k$ and $\hat{\tau}_k$. We know $\hat{G}(t)$ has $\lfloor \alpha t \rfloor + 2$ nodes, $\lfloor \alpha t/2 \rfloor + 1$ of which belong to community 1, and $\lfloor qt \rfloor$ edges, where $q = \frac{1}{2}(q_{with} + q_{bet})$. The corresponding quantities in $G(t)$ are random, but on average, the number of nodes $n(t)$ is close to $\alpha t$, with about $\alpha t/2$ belonging to community 1, and the number of edges $m(t)$ is close to $qt$. The lemmas in this section make these concentrations precise. They rely on Hoeffding's inequality.

**Lemma 20** (Hoeffding's inequality for Bernoulli sums)**.** *Let* $X_1, \ldots, X_n$ *be independent Bernoulli random variables and* $S_n = \sum_{i=1}^n X_i$*. Then*

$$P\left(|S_n - E[S_n]| \geq a\right) \leq 2e^{-2a^2/n}.$$

**Lemma 21.** *Let* $n(t)$ *be the number of nodes in* $G(t)$ *and* $n_i(t)$ *the number of nodes belonging to community* $i$ *in* $G(t)$*,* $i = 1, 2$*. Then for all* $C \geq 1$*,* $t \geq 3$*,*

$$P\left(|n(t) - \alpha t| \geq C\sqrt{t \log t}\right) \leq 2t^{-2} \tag{A.14}$$

$$P\left(|n_i(t) - \frac{1}{2}\alpha t| \geq C\sqrt{t \log t}\right) \leq 2t^{-2}, \qquad i = 1, 2 \tag{A.15}$$

$$E\left[\frac{|n_i(t) - \frac{1}{2}n(t)|}{n(t) + 2}\right] \leq \frac{1}{2\sqrt{\alpha}} t^{-1/2}, \qquad i = 1, 2. \tag{A.16}$$

*Proof.* Since $n(t)$ is binomial with success probability $\alpha$ on $t$ trials, Hoeffding's inequality implies

$$P\left(|n(t) - \alpha t| \geq C\sqrt{t\log t}\right) \leq 2\exp\left(-2(C\sqrt{t\log t})^2/t\right)$$
$$\leq 2\exp\left(-2C^2\log t\right) = 2t^{-2C^2} \leq 2t^{-2}.$$

Similarly for $n_1(t)$ and $n_2(t)$, which are both binomial with success probability $\frac{\alpha}{2}$ on $t$ trials.

Conditional on $n(t) = n$, $n_1(t) \sim Bin(n, \frac{1}{2})$. Therefore

$$E\left[\left|n_1(t) - \frac{1}{2}n(t)\right| \,\middle|\, n(t)\right] \leq \sqrt{\mathrm{Var}(n_1(t)|_{n(t)})} = \sqrt{n(t)/4},$$

and

$$E\left[\frac{|n_i(t) - \frac{1}{2}n(t)|}{n(t)+2}\right] = E\left[\frac{E\left[|n_1(t) - \frac{1}{2}n(t)| \,\middle|\, n(t)\right]}{n(t)+2}\right] \leq E\left[\frac{\sqrt{n(t)}/2}{n(t)+2}\right]$$
$$\leq \frac{1}{2}\sqrt{E\left[\frac{1}{n(t)+1}\right]} = \frac{1}{2}\sqrt{\frac{1-(1-\alpha)^{t+1}}{\alpha(t+1)}} \leq \frac{1}{2\sqrt{\alpha t}}.$$

□

**Lemma 22.** *Let $m(t)$ be the number of edges in $G(t)$ and $m_i(t)$ the number of community $i$ edges in $G(t)$, $i = 1, 2$. For any $C \geq \frac{1+\alpha}{2\alpha}$ and all $t \geq 1$,*

$$P\left(|m(t) - qt| \geq C\sqrt{t\log t}\right) \leq 2t^{-2}. \tag{A.17}$$

*For any $C \geq C_X + 1$, where $C_X$ is defined in (A.3), and all $t \geq 1$,*

$$P\left(\left|m_1(t) - \frac{1}{2}qt\right| \geq C\sqrt{t\log t}\right) \leq 2t^{-2}. \tag{A.18}$$

*Proof.* Define $\xi(t) \in \{0, 1, 2\}$ to be the number of edges added at time $t$, so that

$$m(t) = \sum_{s=1}^{t} \xi(s).$$

Consider the probability that a new edge is added at any time step. From (A.5), we add an edge to community 1 with probability

$$P(X_t = 1) = \frac{m_1(t-1) + \delta n_1(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta} \begin{cases} q_{with} & C(V_t) = 1 \\ q_{bet} & C(V_t) = 2, \end{cases}$$

and from (A.4), we add an edge into community 2 with probability

$$\frac{m_2(t-1) + \delta n_2(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta} \begin{cases} q_{bet} & C(V_t) = 1 \\ q_{with} & C(V_t) = 2. \end{cases}$$

Note that the probability $V_t$ comes from community $i$ is

$$P(C(V_t) = i \mid \mathcal{F}_{t-1}) = \frac{\alpha}{2} + (1-\alpha)\frac{n_i(t-1)+1}{n(t-1)+2}, \qquad i = 1, 2. \tag{A.19}$$

Now, given all information up to time $t-1$, we calculate the expected number of new edges at time $t$ to be

$$\begin{aligned} &E\left[\xi(t) \mid \mathcal{F}_{t-1}\right] \\ &= \frac{m_1(t-1) + \delta n_1(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta}\left[\alpha q + (1-\alpha)\frac{q_{with}(n_1(t-1)+1) + q_{bet}(n_2(t-1)+1)}{n(t-1)+2}\right] \\ &+ \frac{m_2(t-1) + \delta n_2(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta}\left[\alpha q + (1-\alpha)\frac{q_{bet}(n_1(t-1)+1) + q_{with}(n_2(t-1)+1)}{n(t-1)+2}\right]. \end{aligned}$$

In the long run, we expect to add an average of $q$ edges per time step, so we consider

$$\begin{aligned} &E\left[\xi(t) \mid \mathcal{F}_{t-1}\right] - q \\ &= \alpha q + \frac{m_1(t-1) + \delta n_1(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta}(1-\alpha)\frac{(q_{with}+1)n_1(t-1) + q_{bet}(n_2(t-1)+1)}{n(t-1)+2} \\ &+ \frac{m_2(t-1) + \delta n_2(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta}(1-\alpha)\frac{q_{bet}(n_1(t-1)+1) + q_{with}(n_2(t-1)+1)}{n(t-1)+2} - q \end{aligned}$$

$$\begin{aligned} &= (1-\alpha)\frac{m_1(t-1) + \delta n_1(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta}\left(\frac{q_{with}(n_1(t-1)+1) + q_{bet}(n_2(t-1)+1)}{n(t-1)+2} - q\right) \\ &+ (1-\alpha)\frac{m_2(t-1) + \delta n_2(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta}\left(\frac{q_{bet}(n_1(t-1)+1) + q_{with}(n_2(t-1)+1)}{n(t-1)+2} - q\right). \end{aligned}$$

Writing $q = \frac{q_{with}\frac{1}{2}(n(t-1)+2)+q_{bet}\frac{1}{2}(n(t-1)+2)}{n(t-1)+2}$,

$$\begin{aligned}
&\left|E\left[\xi(t) \mid \mathcal{F}_{t-1}\right] - q\right| \\
&\leq (1-\alpha)\frac{m_1(t-1)+\delta n_1(t-1)+\delta}{m(t-1)+\delta n(t-1)+2\delta}\left|q_{with}\frac{n_1(t-1)-\frac{1}{2}n(t-1)}{n(t-1)+2} + q_{bet}\frac{n_2(t-1)-\frac{1}{2}n(t-1)}{n(t-1)+2}\right| \\
&+ (1-\alpha)\frac{m_2(t-1)+\delta n_2(t-1)+\delta}{m(t-1)+\delta n(t-1)+2\delta}\left|q_{bet}\frac{n_1(t-1)-\frac{1}{2}n(t-1)}{n(t-1)+2} + q_{with}\frac{n_2(t-1)-\frac{1}{2}n(t-1)}{n(t-1)+2}\right|.
\end{aligned}$$

Since $n(t-1) = n_1(t-1) + n_2(t-1)$,

$$\begin{aligned}
&\left|E\left[\xi(t) \mid \mathcal{F}_{t-1}\right] - q\right| \\
&\leq (1-\alpha)\frac{m_1(t-1)+\delta n_1(t-1)+\delta}{m(t-1)+\delta n(t-1)+2\delta}\left|q_{with}\frac{n_1(t-1)-\frac{1}{2}n(t-1)}{n(t-1)+2} + q_{bet}\frac{\frac{1}{2}n(t-1)-n_1(t-1)}{n(t-1)+2}\right| \\
&+ (1-\alpha)\frac{m_2(t-1)+\delta n_2(t-1)+\delta}{m(t-1)+\delta n(t-1)+2\delta}\left|q_{bet}\frac{n_1(t-1)-\frac{1}{2}n(t-1)}{n(t-1)+2} + q_{with}\frac{\frac{1}{2}n(t-1)-n_1(t-1)}{n(t-1)+2}\right|
\end{aligned}$$

$$\begin{aligned}
&\leq (1-\alpha)\frac{m_1(t-1)+\delta n_1(t-1)+\delta}{m(t-1)+\delta n(t-1)+2\delta}\left|(q_{with}-q_{bet})\frac{n_1(t-1)-\frac{1}{2}n(t-1)}{n(t-1)+2}\right| \\
&+ (1-\alpha)\frac{m_2(t-1)+\delta n_2(t-1)+\delta}{m(t-1)+\delta n(t-1)+2\delta}\left|(q_{bet}-q_{with})\frac{n_1(t-1)-\frac{1}{2}n(t-1)}{n(t-1)+2}\right|
\end{aligned}$$

$$= (1-\alpha)\left|q_{with}-q_{bet}\right|\frac{\left|n_1(t-1)-\frac{1}{2}n(t-1)\right|}{n(t-1)+2}.$$

Taking another expectation, (A.16) implies

$$|E[\xi(t)] - q| \leq (1-\alpha)|q_{with}-q_{bet}|E\left[\frac{|n_1(t-1)-\frac{1}{2}n(t-1)|}{n(t-1)+2}\right] \leq \frac{(1-\alpha)|q_{with}-q_{bet}|}{2\sqrt{\alpha}}t^{-1/2}.$$

Thus we will write

$$|E[\xi(t)] - q| \leq C_\xi t^{-1/2}, \tag{A.20}$$

where

$$0 < C_\xi = \frac{(1-\alpha)|q_{with}-q_{bet}|}{2\sqrt{\alpha}} \leq \frac{1-\alpha}{2\alpha}.$$

To prove the first part of the lemma, note that $m(t)$ is bounded by a sum of independent trials. Define $S_t^+$ to be the sum of $t$ independent trials, where trial $s$ has probability $q + C_\xi s^{-1/2}$. Then $S_t^+$ stochastically dominates $m(t)$ and

$$E[S_t^+] = \sum_{s=1}^{t} \left(q + C_\xi s^{-1/2}\right) \leq qt + C_\xi \sqrt{t}.$$

Therefore, for any $C \geq \frac{1+\alpha}{2\alpha} \geq C_\xi + 1$,

$$\begin{aligned} P(m(t) \geq qt + C\sqrt{t \log t}) &\leq P(S_t^+ \geq qt + C\sqrt{t \log t}) \\ &\leq P\left(S_t^+ - E[S_t^+] \geq C\sqrt{t \log t} - C_\xi \sqrt{t}\right) \\ &\leq P\left(S_t^+ - E[S_t^+] \geq (C - C_\xi)\sqrt{t \log t}\right). \end{aligned}$$

Then, applying Hoeffding's one-sided inequality,

$$\begin{aligned} P(m(t) \geq qt + C\sqrt{t \log t}) &\leq \exp\left(-2\frac{\left((C - C_\xi)\sqrt{t \log t}\right)^2}{t}\right) \\ &= \exp\left(-2(C - C_\xi)^2 \log t\right) \leq t^{-2}. \end{aligned} \tag{A.21}$$

Similarly, $m(t)$ stochastically dominates a random variable $S_t^-$ defined by the sum of independent trials with success probabilities $q - C_\xi s^{-1/2}$, $s = 1, \ldots, t$. Then we have

$$E[S_t^-] = \sum_{s=1}^{t} \left(q - C_\xi s^{-1/2}\right) \geq qt - C_\xi \sqrt{t}$$

and

$$\begin{aligned} P(m(t) \leq qt - C\sqrt{t \log t}) & \\ &\leq P\left(S_t^- - E[S_t^-] \leq C_\xi \sqrt{t} - C\sqrt{t \log t}\right) \\ &\leq P\left(S_t^- - E[S_t^-] \leq -(C - C_\xi)\sqrt{t \log t}\right) \\ &\leq \exp\left(-2(C - C_\xi)^2 \log t\right) \\ &\leq t^{-2}. \end{aligned} \tag{A.22}$$

Equations (A.21) and (A.22) together prove (A.17).

Now we consider $m_1(t)$. First we note that $m_1(t)$ has half the expectation of $m(t)$. Since $m(t) = m_1(t) + m_2(t)$, and by symmetry of $m_i(t)$, $i = 1, 2$,

$$\begin{aligned}
&E\left[m(t) - qt \,\middle|\, |m(t) - qt| \le \frac{1+\alpha}{2\alpha}\sqrt{t\log t}\right] \\
&= E\left[m_1(t) - \frac{q}{2}t \,\middle|\, |m(t) - qt| \le \frac{1+\alpha}{2\alpha}\sqrt{t\log t}\right] \\
&+ E\left[m_2(t) - \frac{q}{2}t \,\middle|\, |m(t) - qt| \le \frac{1+\alpha}{2\alpha}\sqrt{t\log t}\right] \\
&= 2E\left[m_i(t) - \frac{q}{2}t \,\middle|\, |m(t) - qt| \le \frac{1+\alpha}{2\alpha}\sqrt{t\log t}\right].
\end{aligned}$$

Therefore we have

$$\begin{aligned}
&\left|E\left[m_i(t) - \frac{q}{2}t \,\middle|\, |m(t) - qt| \le \frac{1+\alpha}{2\alpha}\sqrt{t\log t}\right]\right| \\
&= \frac{1}{2}\left|E\left[m(t) - qt \,\middle|\, |m(t) - qt| \le \frac{1+\alpha}{2\alpha}\sqrt{t\log t}\right]\right| \le \frac{1+\alpha}{4\alpha}\sqrt{t\log t} \qquad \text{(A.23)}
\end{aligned}$$

for $t \ge 1$. This will be used later on in the proof of (A.18).

Note that $X_t \in \{0, 1\}$ is the number of community 1 edges added at time $t$, and from (A.5) and (A.19),

$$E\left[X_t \mid \mathcal{F}_{t-1}\right] = \frac{m_1(t-1) + \delta n_1(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta}\left[\alpha q + (1-\alpha)\frac{q_{with}(n_1(t-1)+1) + q_{bet}(n_2(t-1)+1)}{n(t-1)+2}\right].$$

For $t = 1$, we can easily calculate that $E[X_1] = \frac{q}{2}$, since $n_i(0) = m_i(0) = 0$, $i = 1, 2$. For $t \ge 2$, we begin by breaking the right hand side up into two positive factors:

$$\begin{aligned}
\frac{m_1(t-1) + \delta n_1(t-1) + \delta}{m(t-1) + \delta n(t-1) + 2\delta} &= \frac{1}{2} + \frac{m_1(t-1) + \delta n_1(t-1) - \frac{1}{2}(m(t-1) + \delta n(t-1))}{m(t-1) + \delta n(t-1) + 2\delta} \\
&= \frac{1}{2} + \delta\frac{n_1(t-1) - \frac{1}{2}n(t-1)}{m(t-1) + \delta n(t-1) + 2\delta} + \frac{m_1(t-1) - \frac{1}{2}m(t-1)}{m(t-1) + \delta n(t-1) + 2\delta}.
\end{aligned}$$

And since $q = \frac{q_{with}(\frac{1}{2}(n(t-1)+2))+q_{bet}(\frac{1}{2}(n(t-1)+2))}{n(t-1)+2}$ and $n(t-1) = n_1(t-1) + n_2(t-1)$,

$$\begin{aligned}
&\alpha q + (1-\alpha)\frac{q_{with}(n_1(t-1)+1) + q_{bet}(n_2(t-1)+1)}{n(t-1)+2} \\
&\quad = q + (1-\alpha)(q_{with} - q_{bet})\frac{n_1(t-1) - \frac{1}{2}n(t-1)}{n(t-1)+2}.
\end{aligned}$$

Thus

$$\begin{aligned}
E\left[X_t \mid \mathcal{F}_{t-1}\right] &= \left[\frac{1}{2} + \delta\frac{n_1(t-1) - \frac{1}{2}n(t-1)}{m(t-1) + \delta n(t-1) + 2\delta} + \frac{m_1(t-1) - \frac{1}{2}m(t-1)}{m(t-1) + \delta n(t-1) + 2\delta}\right] \\
&\quad \cdot \left[q + (1-\alpha)(q_{with} - q_{bet})\frac{n_1(t-1) - \frac{1}{2}n(t-1)}{n(t-1)+2}\right]
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{q}{2} + \left(\frac{(1-\alpha)|q_{with} - q_{bet}|}{2} + \delta q\right)\frac{|n_1(t-1) - \frac{1}{2}n(t-1)|}{n(t-1)+2} \\
&+ \delta(1-\alpha)|q_{with} - q_{bet}|\left(\frac{|n_1(t-1) - \frac{1}{2}n(t-1)|}{n(t-1)+2}\right)^2 + q\frac{m_1(t-1) - \frac{1}{2}m(t-1)}{m(t-1) + \delta n(t-1) + 2\delta} \\
&+ (1-\alpha)|q_{with} - q_{bet}|\frac{|n_1(t-1) - \frac{1}{2}n(t-1)|}{n(t-1)+2}\frac{|m_1(t-1) - \frac{1}{2}m(t-1)|}{m(t-1)}.
\end{aligned}$$

Since

$$\frac{|n_1(t-1) - \frac{1}{2}n(t-1)|}{n(t-1)+2} \leq 1, \qquad \frac{|m_1(t-1) - \frac{1}{2}m(t-1)|}{m(t-1)} \leq 1$$

at all $t$,

$$\begin{aligned}
E\left[X_t \mid \mathcal{F}_{t-1}\right] - \frac{q}{2} &\leq \left(\delta q + \left(\delta + \frac{3}{2}\right)(1-\alpha)|q_{with} - q_{bet}|\right)\frac{|n_1(t-1) - \frac{1}{2}n(t-1)|}{n(t-1)+2} \\
&\quad + q\frac{m_1(t-1) - \frac{q}{2}(t-1)}{m(t-1) + \delta n(t-1) + 2\delta} + \frac{q}{2}\frac{|m(t-1) - q(t-1)|}{m(t-1) + \delta n(t-1) + 2\delta}. \quad \text{(A.24)}
\end{aligned}$$

Conditional on $|m(t) - qt| \leq \frac{1+\alpha}{2\alpha}\sqrt{t\log t}$, and referring to the definitions of $t^*$ and $\eta$ in (A.1) and (A.2),

$$\begin{aligned}
\frac{t}{m(t) + \delta n(t) + 2\delta} &\leq \frac{t}{2\delta}\mathbb{1}\left(t \leq t^*\right) + \frac{t}{qt - \frac{1+\alpha}{2\alpha}\sqrt{t\log t} + 2\delta}\mathbb{1}\left(t > t^*\right) \\
&\leq \frac{t^*}{2\delta} + \frac{1}{q - \frac{1+\alpha}{2\alpha}\sqrt{\frac{\log t}{t}} + 2\delta}\mathbb{1}\left(t > t^*\right) \leq \frac{1}{\eta} \quad \text{(A.25)}
\end{aligned}$$

for $t \geq 1$.

We will take another expectation on (A.24). Note that

$$
\begin{aligned}
&E\left[\frac{|m(t-1)-q(t-1)|}{m(t-1)+\delta n(t-1)+2\delta}\right] \\
&\leq E\left[\frac{|m(t-1)-q(t-1)|}{m(t-1)+\delta n(t-1)+2\delta}\middle| |m(t-1)-q(t-1)| \leq \frac{1+\alpha}{2\alpha}\sqrt{(t-1)\log(t-1)}\right] \\
&+ P\left(|m(t-1)-q(t-1)| > \frac{1+\alpha}{2\alpha}\sqrt{(t-1)\log(t-1)}\right)
\end{aligned}
$$

For $t \geq 2$, (A.17) and (A.25) imply

$$
\begin{aligned}
&E\left[\frac{|m(t-1)-q(t-1)|}{m(t-1)+\delta n(t-1)+2\delta}\right] \\
&\leq \frac{1+\alpha}{2\alpha}\sqrt{\frac{\log(t-1)}{t-1}}E\left[\frac{t-1}{m(t-1)+\delta n(t-1)+2\delta}\middle| {\scriptstyle |m(t-1)-q(t-1)| \leq \frac{1+\alpha}{2\alpha}\sqrt{(t-1)\log(t-1)}}\right] \\
&+ 2(t-1)^{-2} \\
&\leq \frac{1}{\eta}\frac{1+\alpha}{2\alpha}\sqrt{\frac{\log(t-1)}{t-1}} + 2(t-1)^{-2}. \qquad \text{(A.26)}
\end{aligned}
$$

Also note that

$$
\begin{aligned}
&E\left[\frac{m_1(t-1)-\frac{q}{2}(t-1)}{m(t-1)+\delta n(t-1)+2\delta}\right] \\
&\leq E\left[\frac{m_1(t-1)-\frac{q}{2}(t-1)}{m(t-1)+\delta n(t-1)+2\delta}\middle| |m(t-1)-q(t-1)| \leq \frac{1+\alpha}{2\alpha}\sqrt{(t-1)\log(t-1)}\right] \\
&+ P\left(|m(t-1)-q(t-1)| > \frac{1+\alpha}{2\alpha}\sqrt{(t-1)\log(t-1)}\right)
\end{aligned}
$$

$$
\leq \frac{1}{\eta}E\left[\frac{m_1(t-1)-\frac{q}{2}(t-1)}{t-1}\middle| |m(t-1)-q(t-1)| \leq \frac{1+\alpha}{2\alpha}\sqrt{(t-1)\log(t-1)}\right] + 2(t-1)^{-2}.
$$

Then by (A.23),

$$
E\left[\frac{m_1(t-1)-\frac{q}{2}(t-1)}{m(t-1)+\delta n(t-1)+2\delta}\right] \leq \frac{1}{\eta}\frac{1+\alpha}{4\alpha}\sqrt{\frac{\log(t-1)}{t-1}} + 2(t-1)^{-2}. \qquad \text{(A.27)}
$$

Finally, leveraging (A.16), (A.26), and (A.27), we take another expectation on

(A.24) to see

$$
\begin{aligned}
&E\left[X_t\right] - \frac{q}{2} \\
&\leq \left(\delta q + \left(\delta + \frac{3}{2}\right)(1-\alpha)|q_{with} - q_{bet}|\right)\frac{1}{2\sqrt{\alpha}}(t-1)^{-1/2} \\
&+ q\left(\frac{1}{\eta}\frac{1+\alpha}{4\alpha}\sqrt{\frac{\log(t-1)}{t-1}} + 2(t-1)^{-2}\right) + \frac{q}{2}\left(\frac{1}{\eta}\frac{1+\alpha}{2\alpha}\sqrt{\frac{\log(t-1)}{t-1}} + 2(t-1)^{-2}\right)
\end{aligned}
$$

$$
\leq \frac{\delta q + \left(\delta + \frac{3}{2}\right)(1-\alpha)|q_{with} - q_{bet}|}{2\sqrt{\alpha}}(t-1)^{-1/2} + \frac{q}{\eta}\frac{1+\alpha}{2\alpha}\sqrt{\frac{\log(t-1)}{t-1}} + 3q(t-1)^{-2}. \tag{A.28}
$$

Similarly, bounding from below,

$$
\begin{aligned}
E\left[X_t \mid \mathcal{F}_{t-1}\right] - \frac{q}{2} \geq &-\left(\delta q + \left(\delta + \frac{3}{2}\right)(1-\alpha)|q_{with} - q_{bet}|\right)\frac{|n_1(t-1) - \frac{1}{2}n(t-1)|}{n(t-1)+2} \\
&+ q\frac{m_1(t-1) - \frac{1}{2}q(t-1)}{m(t-1) + \delta n(t-1) + 2\delta} - \frac{q}{2}\frac{|m(t-1) - q(t-1)|}{m(t-1) + \delta n(t-1) + 2\delta}
\end{aligned}
$$

and

$$
\begin{aligned}
&E\left[X_t\right] - \frac{q}{2} \\
&\geq -\frac{\delta q + \left(\delta + \frac{3}{2}\right)(1-\alpha)|q_{with} - q_{bet}|}{2\sqrt{\alpha}}(t-1)^{-1/2} - \frac{q}{\eta}\frac{1+\alpha}{2\alpha}\sqrt{\frac{\log(t-1)}{t-1}} - 3q(t-1)^{-2}.
\end{aligned} \tag{A.29}
$$

Putting (A.28) and (A.29) together, we have

$$
\begin{aligned}
\left|E\left[X_t\right] - \frac{q}{2}\right| &\leq \left(3q + \frac{q}{\eta}\frac{1+\alpha}{2\alpha} + \frac{\delta q + \left(\delta + \frac{3}{2}\right)(1-\alpha)|q_{with} - q_{bet}|}{2\sqrt{\alpha}}\right)\sqrt{\frac{\log(t-1)}{t-1}} \\
&\leq C_X\sqrt{\frac{\log(t-1)}{t-1}},
\end{aligned} \tag{A.30}
$$

where $C_X$ is defined in (A.3).

Now, by (A.30), we can define a random variable $S_1^+$ as the sum of $t$ independent Bernoulli trials, $B_s, s \geq 1$, where for $s \geq 2$, the success probability is $\frac{q}{2} + C_X \sqrt{\frac{\log(s-1)}{s-1}}$. The first trial, $B_1$ has success probability $E[X_1] = \frac{q}{2}$. Now,

$$E[S_1^+] = \frac{q}{2} + \sum_{s=1}^{t-1} \left( \frac{q}{2} + C_X \sqrt{\frac{\log(s)}{s}} \right) \leq \frac{q}{2} t + C_X \sqrt{t \log t}.$$

So for any $C \geq C_X + 1$,

$$\begin{aligned} P\left( m_1(t) - \frac{q}{2} t \geq C \sqrt{t \log t} \right) &\leq P\left( S_1^+ - E[S_1^+] \geq C \sqrt{t \log t} - C_X \sqrt{t \log t} \right) \\ &\leq \exp\left( -2(C - C_X)^2 (t \log t)/t \right) \leq t^{-2}. \end{aligned}$$

Similarly defining $S_1^-$ as the sum of $t$ independent Bernoulli trials with success probabilities $\frac{q}{2} - C_X \sqrt{\frac{\log(s-1)}{s-1}}$, $s \geq 2$, we have

$$P\left( m_1(t) - \frac{q}{2} t \leq -C \sqrt{t \log t} \right) \leq P\left( S_1^- - E[S_1^-] \leq -C \sqrt{t \log t} + C_X \sqrt{t \log t} \right) \leq t^{-2}.$$

□

**Lemma 23.** *Let $\hat{m}_1(t)$ be the number of community 1 edges in $\hat{G}(t)$. There exists $\hat{C}_m \geq 1$ such that*

$$E|\hat{m}_1(t) - \frac{1}{2} qt| \leq \hat{C}_m. \tag{A.31}$$

*Proof.* Let

$$\xi_O(t) = \sum_{s=1}^{t} \mathbb{1}(\lfloor qs \rfloor > \lfloor q(s-1) \rfloor, \lfloor \alpha s \rfloor \text{ is odd })$$

and

$$\xi_E(t) = \sum_{s=1}^{t} \mathbb{1}(\lfloor qs \rfloor > \lfloor q(s-1) \rfloor, \lfloor \alpha s \rfloor \text{ is even}),$$

so that $\xi_O(t) + \xi_E(t) = \lfloor qt \rfloor$. Since $\hat{m}_1(t) = \sum_{s=1}^{t} Y_s$, (A.6) implies

$$E[\hat{m}_1(t)] = \frac{q_{with}}{2q} \xi_O + \frac{q_{bet}}{2q} \xi_E,$$

and noting that $\frac{q_{with}}{2q} + \frac{q_{bet}}{2q} = 1$,

$$E|\hat{m}_1(t) - \frac{1}{2}qt| \leq \frac{q_{with}}{2q}\left|\xi_O(t) - \frac{1}{2}qt\right| + \frac{q_{bet}}{2q}\left|\xi_E(t) - \frac{1}{2}qt\right|.$$

Now, $\xi_O(t)$ and $\xi_E(t)$ are deterministic in $t$, and increase according to a cyclic pattern. Therefore there exists a constant $\hat{C}_m$ such that $|\xi_O(t) - \frac{1}{2}qt|, |\xi_E(t) - \frac{1}{2}qt| \leq \hat{C}_m$, and hence

$$E|\hat{m}_1(t) - \frac{1}{2}qt| \leq \hat{C}_m.$$

□

We have now established that $m_1(t)$ and $\hat{m}_1(t)$ grow approximately linearly in $t$, and stay close to $\frac{q}{2}t$. The relationship between $m_1(t)$ and the time of arrival of the $k$th edge, $\tau_k$, (or $\hat{m}_1(t)$ and $\hat{\tau}_k$) represents the shift from the time scale to indexing by edge arrivals. This inverse relationship suggests that $\tau_k$ and $\hat{\tau}_k$ should look like $\frac{2}{q}k$. The following two lemmas confirm this.

**Lemma 24.** *For each $k \geq 1$ and any $C \geq C_\tau = \frac{2}{q}C_X\sqrt{(\frac{2}{q}+1)\log(\frac{2}{q}+1)}$,*

$$P\left(|\tau_k - \frac{2k}{q}| \geq C\sqrt{k\log k}\right) \leq C_\tau^2 k^{-2}. \tag{A.32}$$

*Proof.* Let's consider the relationship between $\tau_k$ and $m_1(t)$. If $\tau_k > a$ for some $a$, then the time of arrival of the $k$th community 1 edge is after $a$. Thus at time $a$, the graph does not yet have $k$ community 1 edges, and so $m_1(a) < k$.

Similarly, $\tau_k \leq a$ implies $m_1(a) \geq k$. These two facts will give us each side of the inequality in (A.32).

For the first side,

$$P\left(\tau_k > \frac{2}{q}k + C\sqrt{k\log k}\right) \le P\left(\tau_k > \frac{2}{q}k + C_\tau\sqrt{k\log k}\right)$$
$$\le P\left(m_1\left(\frac{2}{q}k + C_\tau\sqrt{k\log k}\right) < k\right)$$
$$= P\left(m_1\left(\frac{2}{q}k + C_\tau\sqrt{k\log k}\right) - \frac{q}{2}\left(\frac{2}{q}k + C_\tau\sqrt{k\log k}\right) < -\frac{q}{2}C_\tau\sqrt{k\log k}\right).$$

It can be seen, after some algebra, that

$$-\frac{q}{2}C_\tau\sqrt{k\log k} \le -C_X\sqrt{(\frac{2}{q}k + C_\tau\sqrt{k})\log(\frac{2}{q}k + C_\tau\sqrt{k})}$$

for all $k \ge C_\tau^2$. Thus

$$P\left(\tau_k > \frac{2}{q}k + C_\tau\sqrt{k\log k}\right)$$
$$\le P\left(m_1\left(\frac{2}{q}k + C_\tau\sqrt{k\log k}\right) - \frac{q}{2}\left(\frac{2}{q}k + C_\tau\sqrt{k\log k}\right) < -C_X\sqrt{\left(\frac{2}{q}k + C_\tau\sqrt{k}\right)\log\left(\frac{2}{q}k + C_\tau\sqrt{k}\right)}\right).$$

Using (A.18), we have,

$$P\left(\tau_k > \frac{2}{q}k + C_\tau\sqrt{k\log k}\right) \le 2\left(\frac{2}{q}k + C_\tau\sqrt{k}\right)^{-2} \le \frac{q^2}{2}k^{-2} \le C_\tau^2 k^{-2}.$$

If $k \le C_\tau^2$, then, trivially,

$$P\left(\tau_k > \frac{2}{q}k + C\sqrt{k\log k}\right) \le 1 \le C_\tau^2 k^{-2}.$$

In the other direction, if we know that $\tau_k \le a$, then we know that there are at least $k$ community 1 edges at time $a$. So

$$P\left(\tau_k \le \frac{2}{q}k - C\sqrt{k\log k}\right)$$
$$\le P\left(\tau_k \le \frac{2}{q}k - C_\tau\sqrt{k\log k}\right)$$
$$\le P\left(m_1\left(\frac{2}{q}k - C_\tau\sqrt{k\log k}\right) \ge k\right)$$
$$= P\left(m_1\left(\frac{2}{q}k - C_\tau\sqrt{k\log k}\right) - \frac{q}{2}\left(\frac{2}{q}k - C_\tau\sqrt{k\log k}\right) \ge \frac{q}{2}C_\tau\sqrt{k\log k}\right)$$
$$\le C_\tau^2 k^{-2}.$$

□

**Lemma 25.** *For each $k \geq 1$ and any $C \geq 1$,*

$$P\left(|\hat{\tau}_k - \frac{2k}{q}| \geq C\sqrt{k\log k}\right) \leq \frac{2\hat{C}_m}{q}k^{-1/2}. \tag{A.33}$$

*Proof.* As for $\tau_k$, $\hat{\tau}_k$ has an inverse relationship with $\hat{m}_1(t)$. If $\hat{\tau}_k > a$, then at time $a$ there are not yet $k$ edges, so $\hat{m}_1(a) < k$. This yields

$$\begin{aligned}
&P\left(\hat{\tau}_k > \frac{2}{q}k + C\sqrt{k\log k}\right) \\
&\quad \leq P\left(\hat{m}_1\left(\frac{2}{q}k + C\sqrt{k\log k}\right) < k\right) \\
&\quad = P\left(\hat{m}_1\left(\frac{2}{q}k + C\sqrt{k\log k}\right) - \frac{q}{2}\left(\frac{2}{q}k + C\sqrt{k\log k}\right) < -\frac{q}{2}C\sqrt{k\log k}\right).
\end{aligned}$$

In the other direction, $\hat{\tau}_k \leq a$ implies $\hat{m}_1(a) \geq k$, so

$$\begin{aligned}
&P\left(\hat{\tau}_k \leq \frac{2}{q}k - C\sqrt{k\log k}\right) \\
&\quad \leq P\left(\hat{m}_1\left(\frac{2}{q}k - C\sqrt{k\log k}\right) \geq k\right) \\
&\quad = P\left(\hat{m}_1\left(\frac{2}{q}k - C\sqrt{k\log k}\right) - \frac{q}{2}\left(\frac{2}{q}k - C\sqrt{k\log k}\right) \geq \frac{q}{2}C\sqrt{k\log k}\right).
\end{aligned}$$

Putting both sides together, and calling on Lemma 23,

$$\begin{aligned}
&P\left(\left|\hat{\tau}_k - \frac{2}{q}k\right| \geq C\sqrt{k\log k}\right) \\
&\quad \leq P\left(\left|\hat{m}_1\left(\frac{2}{q}k - C\sqrt{k\log k}\right) - \frac{q}{2}\left(\frac{2}{q}k - C\sqrt{k\log k}\right)\right| \geq \frac{q}{2}C\sqrt{k\log k}\right) \\
&\quad \leq \frac{E\left|\hat{m}_1\left(\frac{2}{q}k - C\sqrt{k\log k}\right) - \frac{q}{2}\left(\frac{2}{q}k - C\sqrt{k\log k}\right)\right|}{\frac{q}{2}C\sqrt{k\log k}} \\
&\quad \leq \frac{2\hat{C}_m}{q}k^{-1/2}.
\end{aligned}$$

□

## A.5 Proof of Lemma 5

Recall that the bad set $B(k)$ is the set of community 1 nodes belonging to community 1 that have had mismatched degree at any point up to and including the arrival times of the $k$th community 1 edges. The goal of the coupling scheme is to show that the number of nodes in the bad set grows more slowly than linearly, and is therefore negligible in the long run.

**Theorem 21.** *There exists $C_B > 0$ such that*

$$E|B(k)| \leq C_B\left(\sqrt{k \log k} + 1\right)$$

*for all $k$.*

Let us consider the probability that the bad set increases in size due to the addition of the $k$th edges. Define

$$\beta_k = P\left(\text{bad set increases at index } k \mid \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1} \text{do we know } \tau_k???\right).$$

Since

$$E|B(k)| \leq 2 \sum_{j=1}^{k} E[\beta_j],$$

it is sufficient to prove the following lemma, which states that $E[\beta_k] \leq C_\beta \sqrt{\frac{\log(k-1)}{k-1}}$ for $k \geq 2$. Then,

$$\begin{aligned} E|B(k)| &\leq 2\left(1 + 1 + C_\beta \sqrt{\frac{\log 2}{2}} + C_\beta \sqrt{\frac{\log 3}{3}} + \sum_{j=4}^{k-1} C_\beta \sqrt{\frac{\log j}{j}}\right) \\ &\leq 2\left(2 + C_\beta \sqrt{(k-1)\log(k-1)}\right) \\ &\leq C_B \sqrt{k \log k} \end{aligned}$$

for $C_B$ large enough.

**Lemma 26.** *There exists $C_\beta > 0$ such that for any $k \geq 2$,*

$$E[\beta_k] \leq C_\beta \sqrt{\frac{\log(k-1)}{k-1}}.$$

*Proof.* For each $k$, exactly one of the subroutines `Select` $U_k$ `given` $\hat{U}_k$ or `Select` $\hat{U}_k$ `given` $U_k$ is run, either of which can potentially augment the bad set by a maximum of two nodes.

In the case that $\tau_k > \hat{\tau}_k$, $\hat{U}_k$ is selected at time $\hat{\tau}_k$, and the subroutine `Select` $U_k$ `given` $\hat{U}_k$ runs at time $\tau_k$. The bad set increases if any one of the following occurs

1. $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1)$, $\hat{U}_k \in B(k-1)$, and $U_k \in \text{Good}(k)$
2. $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1)$ and $\hat{U}_k \in E(k)$
3. $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(\tau_k - 1)$, $\hat{U}_k \in \text{Good}(k)$, and $U_k \in B(k-1)$
4. $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(\tau_k - 1)$ and $U_k \in E(k)$.

The probability of these occuring is calculated below, using the shorthand notation defined in the subroutine `Select` $U_k$ `given` $\hat{U}_k$.

1. Using the facts that $d_{\text{Good}} \leq y$ and $\hat{d}_B \leq \hat{d}_{B \cup E}$,

$$\begin{aligned}
&P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1),\ \hat{U}_k \in B(k-1),\ U_k \in \text{Good}(k) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&= \frac{\hat{d}_B}{\hat{y}} \sum_{w \in \text{Good}(k)} \frac{d_w}{\hat{d}_{B \cup E}} \left(\frac{\hat{y}}{y} - 1\right) P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&\leq \frac{d_{\text{Good}}}{y} \left(\frac{\hat{y} - y}{\hat{y}}\right) P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&\leq \frac{|\hat{y} - y|}{\hat{y}} P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&\leq \frac{|\hat{y} - y|}{y} P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right).
\end{aligned}$$

The last step holds since $\hat{y} \geq y$ when $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1)$.

2. Note that any node has zero in-degree, so $d_w + \delta$ for $w \in E(k)$. Since $E(k) \subset \{\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \wedge n_1(\tau_k - 1), \ldots, \lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \vee n_1(\tau_k - 1)\}$, then

$$d_E = \hat{d}_E = \delta |E(k)| \leq \delta |\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor - n_1(\tau_k - 1)| = |\hat{y} - y|. \tag{A.34}$$

Thus

$$\begin{aligned}
&P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1),\ \hat{U}_k \in E(k) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&= \frac{\hat{d}_E}{\hat{y}} P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&\leq \frac{|\hat{y} - y|}{\hat{y}} P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&\leq \frac{|\hat{y} - y|}{y} P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor > n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right).
\end{aligned}$$

3. Using that $\hat{d}_{\text{Good}} \leq \hat{y}$ and $d_B \leq d_{B \cup E}$,

$$\begin{aligned}
&P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(\tau_k - 1),\ \hat{U}_k \in \text{Good}(k),\ U_k \in B(k-1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&\leq \frac{\hat{d}_{\text{Good}}}{\hat{y}} \sum_{w \in B(k-1)} \frac{d_w}{d_{B \cup E}} \left(1 - \frac{\hat{y}}{y}\right) P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&\leq \frac{d_B}{d_{B \cup E}} \left(1 - \frac{\hat{y}}{y}\right) P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&\leq \frac{|\hat{y} - y|}{y} P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&\leq \frac{|\hat{y} - y|}{\hat{y}} P\left(\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right),
\end{aligned}$$

since $\hat{y} \leq y$ when $\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor \leq n_1(\tau_k - 1)$.

4. Using (A.34) and Lemma 19,

$$\begin{aligned}
&P\left(\lfloor\alpha(\hat{\tau}_k - 1)/2\rfloor \le n_1(\tau_k - 1),\ U_k \in E(k) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&= \left[\sum_{w \in E(k)} \frac{d_w}{y}\right] P\left(\lfloor\alpha(\hat{\tau}_k - 1)/2\rfloor \le n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&= \frac{d_E}{y} P\left(\lfloor\alpha(\hat{\tau}_k - 1)/2\rfloor \le n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&\le \frac{|\hat{y} - y|}{y} P\left(\lfloor\alpha(\hat{\tau}_k - 1)/2\rfloor \le n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \\
&\le \frac{|\hat{y} - y|}{\hat{y}} P\left(\lfloor\alpha(\hat{\tau}_k - 1)/2\rfloor \le n_1(\tau_k - 1) \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right),
\end{aligned}$$

since $\hat{y} \le y$ when $\lfloor\alpha(\hat{\tau}_k - 1)/2\rfloor \le n_1(\tau_k - 1)$.

Combining all four calculations, we have both of the following statements:

$$P\left(\text{bad set increases at index } k \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \le 2\frac{|\hat{y} - y|}{\hat{y}}$$

and

$$P\left(\text{bad set increases at index } k \mid \tau_k > \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \le 2\frac{|\hat{y} - y|}{y}.$$

By a symmetric argument, when $\tau_k \le \hat{\tau}_k$, we have

$$P\left(\text{bad set increases at index } k \mid \tau_k \le \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \le 2\frac{|\hat{y} - y|}{\hat{y}}$$

and

$$P\left(\text{bad set increases at index } k \mid \tau_k \le \hat{\tau}_k, \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \le 2\frac{|\hat{y} - y|}{y}.$$

Therefore,

$$\begin{aligned}
\beta_k &= P\left(\text{bad set increases at index } k \mid \mathcal{F}_{\tau_k - 1}, \hat{\mathcal{F}}_{\hat{\tau}_k - 1}\right) \le 2\frac{|\hat{y} - y|}{\hat{y}} \\
&= 2\delta\frac{|\lfloor\alpha(\hat{\tau}_k - 1)/2\rfloor - n_1(\tau_k - 1)|}{k - 1 + \delta\lfloor\alpha(\hat{\tau}_k - 1)/2\rfloor + \delta} \le 2\delta\frac{|\lfloor\alpha(\hat{\tau}_k - 1)/2\rfloor - n_1(\tau_k - 1)|}{k - 1 + \delta(\alpha(\hat{\tau}_k - 1)/2 - 1) + \delta} \\
&\le 2\delta\frac{|\lfloor\alpha(\hat{\tau}_k - 1)/2\rfloor - n_1(\tau_k - 1)|}{k - 1 + \frac{\alpha\delta}{2}(\hat{\tau}_k - 1)}
\end{aligned} \tag{A.35}$$

Since $\hat{\tau}_k \geq k$, we have, for $k \geq 2$,

$$\begin{aligned}\beta_k &\leq \frac{4}{\alpha}\frac{|\lfloor \alpha(\hat{\tau}_k - 1)/2 \rfloor - n_1(\hat{\tau}_k - 1)|}{\hat{\tau}_k - 1} + \frac{4}{\alpha}\frac{|n_1(\hat{\tau}_k - 1) - n_1(\tau_k - 1)|}{\hat{\tau}_k - 1} \\ &\leq \frac{4}{\alpha}\left[\frac{1}{k-1} + \frac{|\alpha(\tau_k - 1)/2 - n_1(\hat{\tau}_k - 1)|}{\hat{\tau}_k - 1} + \frac{|\hat{\tau}_k - \tau_k|}{k-1}\right] \end{aligned} \tag{A.36}$$

From (A.15), we know

$$P\left(|\alpha(\hat{\tau}_k - 1)/2 - n_1(\hat{\tau}_k - 1)| \geq 2\sqrt{(\hat{\tau}_k - 1)\log(\hat{\tau}_k - 1)}\right) \leq 2(\hat{\tau}_k - 1)^{-2} \leq 2(k-1)^{-2}, \tag{A.37}$$

and by Lemmas 24 and 25, noting that $C_\tau \geq 1$,

$$\begin{aligned} &P(|\hat{\tau}_k - \tau_k| \geq 2C_\tau\sqrt{k\log k}) \\ &\quad \leq P\left(|\hat{\tau}_k - \frac{2k}{q}| \geq C_\tau\sqrt{k\log k}\right) + P\left(|\tau_k - \frac{2k}{q}| \geq C_\tau\sqrt{k\log k}\right) \\ &\quad \leq \frac{2\hat{C}_m}{q}k^{-1/2} + C_\tau^2 k^{-2}. \end{aligned} \tag{A.38}$$

Putting (A.36), (A.37), and (A.38) together,

$$\begin{aligned} E\beta_k &\leq \frac{4}{\alpha}\left[\frac{1}{k-1} + 2\sqrt{\frac{\log(\hat{\tau}_k - 1)}{\hat{\tau}_k - 1}} + 2C_\tau\frac{\sqrt{k\log k}}{k-1}\right] \\ &\quad + P\left(|\alpha(\hat{\tau}_k - 1)/2 - n_1(\hat{\tau}_k - 1)| \geq 2\sqrt{(\hat{\tau}_k - 1)\log(\hat{\tau}_k - 1)}\right) \\ &\quad + P\left(|\hat{\tau}_k - \tau_k| \geq 2C_\tau\sqrt{k\log k}\right) \\ &\leq \frac{4}{\alpha}\left[\frac{1}{k-1} + (2 + 3C_\tau)\sqrt{\frac{\log(k-1)}{k-1}}\right] + 2(k-1)^{-2} + \frac{2\hat{C}_m}{q}k^{-1/2} + C_\tau^2 k^{-2} \\ &\leq C_\beta\sqrt{\frac{\log(k-1)}{k-1}}, \end{aligned}$$

where

$$C_\beta = \frac{12}{\alpha}(1 + C_\tau) + 2 + \frac{2\hat{C}_m}{q} + C_\tau^2.$$

□

*Proof of Lemma 5.* We use

$$\max_{i,j} \left| N_t(i,j) - \hat{N}_t(i,j) \right| \leq \sum_{i,j} \left| N_t(i,j) - \hat{N}_t(i,j) \right|$$
$$= |\lfloor \alpha t/2 \rfloor - n_1(t)| + \sum_{v=1}^{\lfloor \alpha t/2 \rfloor \wedge n_1(t)} \mathbb{1}(D^v(t) \neq \hat{D}^v(t)). \tag{A.39}$$

The summation of indicators counts the number of community 1 nodes that exist in both graphs that do not have matching degree at time $t$. To relate this sum (at time $t$) to the size of the bad set, we must look back to the "most recent" $k$. Let $m_1(t)$ denote the number of edges into community 1 of $G(t)$ and $\hat{m}_1(t)$ the number of edges into community 1 of $\hat{G}(t)$. Set

$$k(t) = \min\{m_1(t), \hat{m}_1(t)\}.$$

Clearly, any node in the bad set $B(k(t))$ may not have identical in-degree at time $t$.

Any contributions to the sum which are not in the bad set at index $k(t)$ must result from any community 1 edges added after the $k(t)$th edge is added, but before time $t$. There are $|m_1(t) - \hat{m}_1(t)|$ additional edges, each of which can cause at most one new mismatch.

Thus we can write

$$\max_{i,j} \left| N_t(i,j) - \hat{N}_t(i,j) \right| \leq |\lfloor \alpha t/2 \rfloor - n_1(t)| + \sum_{v \in C_1} \mathbb{1}(D^v(t) \neq \hat{D}^v(t))$$
$$\leq |\lfloor \alpha t/2 \rfloor - n_1(t)| + |B(k(t))| + |m_1(t) - \hat{m}_1(t)|.$$

Using (A.15), (A.18), Lemma 23, Theorem 21, and the fact that $k(t) \leq t$, we

have, for any $C \geq 4(C_X + 1)$,

$$
\begin{aligned}
&P\left(\max_{i,j} \left|N_t(i,j) - \hat{N}_t(i,j)\right| \geq Ct^{3/5}\right) \\
&\leq P\left(|\lfloor \alpha t/2 \rfloor - n_1(t)| + |B(k(t))| + |m_1(t) - \hat{m}_1(t)| \geq Ct^{3/5}\right) \\
&\leq P\left(|\lfloor \alpha t/2 \rfloor - n_1(t)| \geq \frac{C}{4} t^{3/5}\right) + \frac{E|B(k(t))|}{\frac{C}{4}t^{3/5}} \\
&+ P\left(|m_1(t) - \frac{1}{2}qt| \geq \frac{C}{4} t^{3/5}\right) + P\left(|\hat{m}_1(t) - \frac{1}{2}qt| \geq \frac{C}{4} t^{3/5}\right) \\
&\leq 2t^{-2} + \frac{C_B\left(\sqrt{k(t)\log(k(t))} + 1\right)}{\frac{C}{4}t^{3/5}} + P\left(|m_1(t) - \frac{1}{2}qt| \geq \frac{C}{4}\sqrt{t\log t}\right) + \frac{E|\hat{m}_1(t) - \frac{1}{2}qt|}{\frac{C}{4}t^{3/5}} \\
&\leq 2t^{-2} + \frac{C_B\left(\sqrt{t\log t} + 1\right)}{\frac{C}{4}t^{3/5}} + 2t^{-2} + \frac{\hat{C}_m}{\frac{C}{4}t^{3/5}} \\
&\leq C_N \frac{\sqrt{\log t}}{t^{1/10}} = o(1),
\end{aligned}
$$

where

$$C_N = 4 + \frac{C_B + \hat{C}_m}{C_X + 1}.$$

□

APPENDIX B

## PROOF OF EXPECTED DEGREE SEQUENCE

Towards proving Lemma 7, we define

$$\epsilon_t(i,j) = E[\hat{N}_t(i,j)] - \frac{\alpha}{2} t p(i,j),$$

and we will show that $|\epsilon_t(i,j)| \le C_\epsilon(t^{3/5}+1)$ via induction.

Recall the indicator functions

$$\mathbb{1}_{\alpha,t} = \mathbb{1}(\lfloor \alpha(t+1) \rfloor > \lfloor \alpha t \rfloor)$$

$$\mathbb{1}_{q,t} = \mathbb{1}(\lfloor q(t+1) \rfloor > \lfloor q t \rfloor)$$

$$\mathbb{1}_{\text{odd},t} = \mathbb{1}(\lfloor \alpha(t+1) \rfloor \text{ is odd})$$

and the one-step recursion for $E[\hat{N}_{t+1}(i,j)]$:

$$\begin{aligned} E[\hat{N}_{t+1}(i,j)] &= \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t}\mathbb{1}(i=j=0) + \mathbb{1}_{q,t}E[\Delta_t(i,j)] \\ &+ \left(1 - \mathbb{1}_{q,t}\frac{i+j+\delta}{\frac{1}{2}(q+\alpha\delta)t}\left(\frac{q_{with}}{2q}\mathbb{1}_{\text{odd},t} + \frac{q_{bet}}{2q}(1-\mathbb{1}_{\text{odd},t})\right)\right)E[\hat{N}_t(i,j)] \\ &+ \mathbb{1}_{q,t}\frac{i+j-1+\delta}{\frac{1}{2}(q+\alpha\delta)t}\left(\frac{q_{with}}{2q}\mathbb{1}_{\text{odd},t}E[\hat{N}_t(i-1,j)] + \frac{q_{bet}}{2q}(1-\mathbb{1}_{\text{odd},t})E[\hat{N}_t(i,j-1)]\right), \end{aligned} \tag{B.1}$$

where $E|\Delta_t(i,j)| \le \frac{T+\delta}{\frac{1}{2}(q+\alpha\delta)} t^{-1}$ by Lemma 9.

Recall that by Assumption 2,

$$q_{with} \le q + \alpha\delta, \qquad q_{bet} \le q + \alpha\delta, \tag{B.2}$$

$K$ is defined as

$$K = \min\left\{k \ge 0 : kT \ge \max\left\{4,\ \frac{\delta}{q}\left(\frac{q_{with}}{q+\alpha\delta-q_{with}}\right),\ \frac{\delta}{q}\left(\frac{q_{bet}}{q+\alpha\delta-q_{bet}}\right)\right\}\right\}, \tag{B.3}$$

and $C_\epsilon$ is chosen to satisfy

$$C_\epsilon \ge \max\left\{T\left(1+\frac{\alpha}{2}+\frac{3\alpha}{q}\right),\ KT,\ \left(\frac{2q}{q_{with}}+\frac{2q}{q_{bet}}\right)\left(\frac{T+\delta}{(KT)^{3/5}-2}\right)\right\}. \tag{B.4}$$

*Proof of Lemma 7.* Since $q$ is rational, there exists a period $T$ such that for all $t \geq 0$,

$$\mathbb{1}_{q,t} = \mathbb{1}_{q,t \bmod T}$$

$$\mathbb{1}_{\alpha,t} = \mathbb{1}_{\alpha,t \bmod T}$$

$$\mathbb{1}_{\text{odd},t} = \mathbb{1}_{\text{odd},t \bmod T}$$

and for any nonnegative integer $k$,

$$\sum_{t=kT}^{kT+T-1} \mathbb{1}_{q,t} = \sum_{t=0}^{T-1} \mathbb{1}_{q,t} = qT \tag{B.5}$$

and the identities

$$\sum_{t=0}^{T-1} \mathbb{1}_{q,t}\mathbb{1}_{\text{odd},t} = \sum_{t=0}^{T-1} \mathbb{1}_{q,t}(1 - \mathbb{1}_{\text{odd},t}) = \frac{qT}{2} \tag{B.6}$$

$$\sum_{t=0}^{T-1} \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t} = \sum_{t=0}^{T-1} \mathbb{1}_{\alpha,t}(1 - \mathbb{1}_{\text{odd},t}) = \frac{\alpha T}{2} \tag{B.7}$$

hold.

For small $t$, the result $|\epsilon_t(i,j)| \leq C_\epsilon(t^{3/5} + 1)$ holds trivially. Specifically, for $t = 0$, note that

$$|\epsilon_0(i,j)| = \left|E[\hat{N}_0(i,j)]\right| = \mathbb{1}_{i=j=0} \leq C_\epsilon,$$

and for $1 \leq t \leq KT$,

$$|\epsilon_t(i,j)| \leq t \leq C_\epsilon \leq C_\epsilon t^{3/5} \tag{B.8}$$

by the definition of $C_\epsilon$. For $t \geq KT$ we will use the induction hypotheses

$$\epsilon_t(i,j) \leq C_\epsilon t^{3/5} + f_t(i,j) \tag{B.9}$$

$$\epsilon_t(i,j) \geq -C_\epsilon t^{3/5} + f_t(i,j), \tag{B.10}$$

where

$$\begin{aligned} f_t(i,j) = \sum_{s=0}^{(t \bmod T)+1} \Bigg( & \mathbb{1}_{\alpha,s}\mathbb{1}_{\text{odd},s}\mathbb{1}_{i=j=0} - \frac{2}{q}\frac{\alpha}{2}\mathbb{1}_{q,s}\mathbb{1}_{i=j=0} - \frac{\alpha}{2}p(i,j) \\ & + \frac{2}{q}\frac{\alpha}{2}p(i,j)\mathbb{1}_{q,s} - \frac{2}{q}\gamma_w(i,j)\mathbb{1}_{q,s}(1 - \mathbb{1}_{\text{odd},s}) - \frac{2}{q}\gamma_b(i,j)\mathbb{1}_{q,s}\mathbb{1}_{\text{odd},s}\Bigg) \end{aligned} \tag{B.11}$$

if $t \bmod T \geq 1$, and $f_t(i,j) = 0$ if $t \bmod T = 0$. Since $(i+j+\delta)p(i,j) \leq 1$ for all $i, j \geq 0$, (B.11) implies

$$
\begin{aligned}
&(i+j+\delta)|f_t(i,j)| \qquad\qquad (B.12)\\
&= (i+j+\delta)\left|\sum_{s=0}^{(t \bmod T)-1}\left(\mathbb{1}_{\alpha,s}\mathbb{1}_{\text{odd},s}\mathbb{1}_{i=j=0} - \frac{2}{q}\frac{\alpha}{2}\mathbb{1}_{q,s}\mathbb{1}_{i=j=0} - \frac{\alpha}{2}p(i,j)\right.\right.\\
&\left.\left.+\frac{2}{q}\frac{\alpha}{2}p(i,j)\mathbb{1}_{q,s} - \frac{2}{q}\gamma_w(i,j)\mathbb{1}_{q,s}(1-\mathbb{1}_{\text{odd},s}) - \frac{2}{q}\gamma_b(i,j)\mathbb{1}_{q,s}\mathbb{1}_{\text{odd},s}\right)\right|\\
&\leq (i+j+\delta)T\left(1+\frac{\alpha}{q}+\left(1+\frac{2}{q}\right)\frac{\alpha}{2}p(i,j)+\frac{2}{q}\left|\gamma_w(i,j)+\gamma_b(i,j)\right|\right)\\
&\leq T\left[1+\frac{\alpha}{q}+\left(1+\frac{4}{q}\right)\frac{\alpha}{2}(i+j+\delta)p(i,j)\right] \leq T\left(1+\frac{\alpha}{2}+\frac{3\alpha}{q}\right) \leq C_\epsilon.
\end{aligned}
$$

Thus (B.9) and (B.10) are sufficient to prove that $|\epsilon_t(i,j)| \leq C_\epsilon(t^{3/5}+1)$ for all $t$.

Taking $t = KT$ in (B.8) establishes the base case for (B.9), and we now assume that (B.9) holds for some $t \geq KT$, for all $i, j \geq 0$. We will show that it also holds for $t+1$.

We begin by noting that

$$
\begin{aligned}
\frac{\alpha}{2}(t+1)p(i,j) &= \frac{\alpha}{2}tp(i,j) + \frac{\alpha}{2}p(i,j)\\
&= \frac{\alpha}{2}tp(i,j) + \left[1-\frac{2}{q}\right]\frac{\alpha}{2}p(i,j) + \frac{2}{q}\frac{\alpha}{2}p(i,j). \qquad (B.13)
\end{aligned}
$$

Defining

$$
\gamma_w(i,j) = \frac{q_{with}}{2q}\frac{\alpha}{2}\left(\frac{i+j-1+\delta}{1+c}p(i-1,j) - \frac{i+j+\delta}{1+c}p(i,j)\right)
$$

and

$$
\gamma_b(i,j) = \frac{q_{bet}}{2q}\frac{\alpha}{2}\left(\frac{i+j-1+\delta}{1+c}p(i,j-1) - \frac{i+j+\delta}{1+c}p(i,j)\right),
$$

and referring to (3.4), we rewrite (B.13) as

$$
\begin{aligned}
&\frac{\alpha}{2}(t+1)p(i,j) \qquad\qquad (B.14)\\
&= \frac{\alpha}{2}tp(i,j) + \left[1-\frac{2}{q}\right]\frac{\alpha}{2}p(i,j) + \frac{2}{q}\left(\frac{\alpha}{2}\mathbb{1}_{i=j=0} + \gamma_w(i,j) + \gamma_b(i,j)\right).
\end{aligned}
$$

At this point we will consider $i, j$ such that $i + j > 0$, and assume that the case $i = j = 0$ holds. We will return to this special case later on. We now show that (B.9) holds for $t + 1$ by dividing into three cases: (a) $\mathbb{1}_{q,t} = 0$, (b) $\mathbb{1}_{q,t} = \mathbb{1}_{\text{odd},t} = 1$, and (c) $\mathbb{1}_{q,t} = 1 - \mathbb{1}_{\text{odd},t} = 1$.

**Case (a):** $\mathbb{1}_{q,t} = 0$.

From (B.1), we see that

$$\begin{aligned}
\epsilon_{t+1}(i,j) &= E[\hat{N}_{t+1}(i,j)] - \frac{\alpha}{2}(t+1)p(i,j) \\
&= \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t}\mathbb{1}_{i=j=0} + E[\hat{N}_t(i,j)] - \frac{\alpha}{2}tp(i,j) - \frac{\alpha}{2}p(i,j) \\
&= \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t}\mathbb{1}_{i=j=0} + \epsilon_t(i,j) - \frac{\alpha}{2}p(i,j).
\end{aligned}$$

Since $i + j > 0$, and using (B.9),

$$\epsilon_{t+1}(i,j) = \epsilon_t(i,j) - \frac{\alpha}{2}p(i,j) \le C_\epsilon t^{3/5} + f_t(i,j) - \frac{\alpha}{2}p(i,j). \tag{B.15}$$

**Case (b):** $\mathbb{1}_{q,t} = \mathbb{1}_{\text{odd},t} = 1$.

From (B.1) and (B.14),

$$\begin{aligned}
\epsilon_{t+1}(i,j) &= E[\hat{N}_{t+1}(i,j)] - \frac{\alpha}{2}(t+1)p(i,j) \\
&= \left(1 - \frac{q_{with}}{2q}\frac{i+j+\delta}{\frac{1}{2}(q+\alpha\delta)t}\right)E[\hat{N}_t(i,j)] + \frac{i+j-1+\delta}{\frac{1}{2}(q+\alpha\delta)t}\frac{q_{with}}{2q}E[\hat{N}_t(i-1,j)] \\
&+ E[\Delta_t(i,j)] - \frac{\alpha}{2}tp(i,j) + \left(\frac{2}{q} - 1\right)\frac{\alpha}{2}p(i,j) - \frac{2}{q}\left(\gamma_w(i,j) + \gamma_b(i,j)\right) \\
&= \left(1 - \frac{q_{with}}{2q}\frac{i+j+\delta}{\frac{1}{2}(q+\alpha\delta)t}\right)\epsilon_t(i,j) + \frac{i+j-1+\delta}{\frac{1}{2}(q+\alpha\delta)t}\frac{q_{with}}{2q}\epsilon_t(i-1,j) \\
&+ E[\Delta_t(i,j)] + \left(\frac{2}{q} - 1\right)\frac{\alpha}{2}p(i,j) - \frac{2}{q}\gamma_b(i,j).
\end{aligned}$$

Note that at time $t$, there are a total of $\lfloor qt \rfloor$ edges in the graph, so for all $i, j$ such that $\hat{N}_t(i,j) > 0$, we must have $i + j < \lfloor qt \rfloor$. Then (B.2) and (B.3) imply that for all

$t \geq KT$, we have $\frac{q_{with}}{2q} \frac{i+j+\delta}{\frac{1}{2}(q+\alpha\delta)t} \leq 1$ for eacg $i, j$ such that $\hat{N}_t(i,j) > 0$. This means that we can use the induction hypothesis (B.9) to see

$$\begin{aligned}\epsilon_{t+1}(i,j) \leq & \left(1 - \frac{q_{with}}{2q} \frac{i+j+\delta}{\frac{1}{2}(q+\alpha\delta)t}\right)\left(C_\epsilon t^{3/5} + f_t(i,j)\right) \\ & + \frac{i+j-1+\delta}{\frac{1}{2}(q+\alpha\delta)t} \frac{q_{with}}{2q}\left(C_\epsilon t^{3/5} + f_t(i-1,j)\right) \\ & + E[\Delta_t(i,j)] + \left(\frac{2}{q} - 1\right)\frac{\alpha}{2}p(i,j) - \frac{2}{q}\gamma_b(i,j).\end{aligned}$$

By (B.12) and Lemma 9,

$$\begin{aligned}\epsilon_{t+1}(i,j) \leq & \left(1 - \frac{q_{with}}{2q} \frac{1}{\frac{1}{2}(q+\alpha\delta)t}\right) C_\epsilon t^{3/5} + 2\frac{q_{with}}{2q} \frac{1}{\frac{1}{2}(q+\alpha\delta)t} C_\epsilon \\ & + \frac{T+\delta}{\frac{1}{2}(q+\alpha\delta)t} + \left(\frac{2}{q} - 1\right)\frac{\alpha}{2}p(i,j) - \frac{2}{q}\gamma_b(i,j) + f_t(i,j).\end{aligned} \tag{B.16}$$

Since $t \geq KT$ and $C_\epsilon \geq \frac{2q}{q_{with}} \frac{T+\delta}{(KT)^{3/5}-2}$ by (B.4),

$$\epsilon_{t+1}(i,j) \leq C_\epsilon (t+1)^{3/5} + \left(\frac{2}{q} - 1\right)\frac{\alpha}{2}p(i,j) - \frac{2}{q}\gamma_b(i,j) + f_t(i,j). \tag{B.17}$$

**Case (c):** $\mathbb{1}_{q,t} = 1 - \mathbb{1}_{\text{odd},t} = 1$.

From (B.1) and (B.14),

$$\epsilon_{t+1}(i,j) = E[\hat{N}_{t+1}(i,j)] - \frac{\alpha}{2}(t+1)p(i,j)$$

$$\begin{aligned}= & \left(1 - \frac{q_{bet}}{2q} \frac{i+j+\delta}{(q+\alpha\delta)t}\right) E[\hat{N}_t(i,j)] + \frac{i+j-1+\delta}{(q+\alpha\delta)t} \frac{q_{bet}}{2q} E[\hat{N}_t(i,j-1)] \\ & + E[\Delta_t(i,j)] - \frac{\alpha}{2}tp(i,j) + \left(\frac{2}{q} - 1\right)\frac{\alpha}{2}p(i,j) - \frac{2}{q}\left(\gamma_w(i,j) + \gamma_b(i,j)\right)\end{aligned}$$

$$\begin{aligned}= & \left(1 - \frac{q_{bet}}{2q} \frac{i+j+\delta}{(q+\alpha\delta)t}\right) \epsilon_t(i,j) + \frac{i+j-1+\delta}{(q+\alpha\delta)t} \frac{q_{bet}}{2q} \epsilon_t(i,j-1) \\ & + E[\Delta_t(i,j)] + \left(\frac{2}{q} - 1\right)\frac{\alpha}{2}p(i,j) - \frac{2}{q}\gamma_w(i,j).\end{aligned}$$

As in Case (b), (B.2) and (B.3) imply that for all $i, j$ such that $\hat{N}_t(i,j) > 0$, we have $\frac{q_{bet}}{2q}\frac{i+j+\delta}{\frac{1}{2}(q_\alpha\delta)t} \leq 1$. Then by (B.9),

$$\begin{aligned}\epsilon_{t+1}(i,j) \leq &\left(1 - \frac{q_{bet}}{2q}\frac{i+j+\delta}{\frac{1}{2}(q+\alpha\delta)t}\right)\left(C_\epsilon t^{3/5} + f_t(i,j)\right)\\ &+ \frac{i+j-1+\delta}{\frac{1}{2}(q+\alpha\delta)t}\frac{q_{bet}}{2q}\left(C_\epsilon t^{3/5} + f_t(i,j-1)\right)\\ &+ E[\Delta_t(i,j)] + \left(\frac{2}{q} - 1\right)\frac{\alpha}{2}p(i,j) - \frac{2}{q}\gamma_w(i,j).\end{aligned}$$

By (B.12) and Lemma 9,

$$\epsilon_{t+1}(i,j) \leq C_\epsilon(t+1)^{3/5} + f_t(i,j) + \left(\frac{2}{q} - 1\right)\frac{\alpha}{2}p(i,j) - \frac{2}{q}\gamma_w(i,j). \tag{B.18}$$

At this point, we have shown, through (B.15), (B.17), and (B.18), that

$$\begin{aligned}\epsilon_{t+1} \leq C_\epsilon(t+1)^{3/5} + f_t(i,j) - \frac{\alpha}{2}p(i,j) + \frac{2}{q}\frac{\alpha}{2}p(i,j)\mathbb{1}_{q,t}\\ - \frac{2}{q}\gamma_w(i,j)\mathbb{1}_{q,t}(1 - \mathbb{1}_{\text{odd},t}) - \frac{2}{q}\gamma_b(i,j)\mathbb{1}_{q,t}\mathbb{1}_{\text{odd},t}.\end{aligned} \tag{B.19}$$

Now we must show that the terms to the right of and including $f_t(i,j)$ in (B.19) sum to $f_{t+1}(i,j)$. Recalling that the form of $f_{t+1}(i,j)$ depends on whether $t+1$ is a multiple of $T$,

$$\begin{aligned}&f_t(i,j) - \frac{\alpha}{2}p(i,j) + \frac{2}{q}\frac{\alpha}{2}p(i,j)\mathbb{1}_{q,t} - \frac{2}{q}\gamma_w(i,j)\mathbb{1}_{q,t}(1 - \mathbb{1}_{\text{odd},t}) - \frac{2}{q}\gamma_b(i,j)\mathbb{1}_{q,t}\mathbb{1}_{\text{odd},t}\\ &= \sum_{s=0}^{(t \bmod T)-1}\left(-\frac{\alpha}{2}p(i,j) + \frac{2}{q}\frac{\alpha}{2}p(i,j)\mathbb{1}_{q,s} - \frac{2}{q}\gamma_w(i,j)\mathbb{1}_{q,s}(1 - \mathbb{1}_{\text{odd},s})\right.\\ &\left.-\frac{2}{q}\gamma_b(i,j)\mathbb{1}_{q,s}\mathbb{1}_{\text{odd},s}\right) - \frac{\alpha}{2}p(i,j) + \frac{2}{q}\frac{\alpha}{2}p(i,j)\mathbb{1}_{q,t \bmod T}\\ &- \frac{2}{q}\gamma_w(i,j)\mathbb{1}_{q,t \bmod T}(1 - \mathbb{1}_{\text{odd},t \bmod T}) - \frac{2}{q}\gamma_b(i,j)\mathbb{1}_{q,t \bmod T}\mathbb{1}_{\text{odd},t \bmod T}\\ &= \sum_{s=0}^{t \bmod T}\left(-\frac{\alpha}{2}p(i,j) + \frac{2}{q}\frac{\alpha}{2}p(i,j)\mathbb{1}_{q,s} - \frac{2}{q}\gamma_w(i,j)\mathbb{1}_{q,s}(1 - \mathbb{1}_{\text{odd},s})\right.\\ &\left.-\frac{2}{q}\gamma_b(i,j)\mathbb{1}_{q,s}\mathbb{1}_{\text{odd},s}\right).\end{aligned} \tag{B.20}$$

If $(t+1) \bmod T > 0$, then $t \bmod T = ((t+1) \bmod T) - 1$, and so (B.20) is simply the definition of $f_{t+1}(i,j)$ for $i + j > 0$.

If, on the other hand, $(t+1) \bmod T = 0$, then $t \bmod T = T-1$, so (B.20) becomes

$$\begin{aligned}
&= \sum_{s=0}^{T-1} \left( -\frac{\alpha}{2} p(i,j) + \frac{2}{q}\frac{\alpha}{2} p(i,j) \mathbb{1}_{q,s} - \frac{2}{q}\gamma_w(i,j)\mathbb{1}_{q,s}(1 - \mathbb{1}_{\text{odd},s}) - \frac{2}{q}\gamma_b(i,j)\mathbb{1}_{q,s}\mathbb{1}_{\text{odd},s} \right) \\
&= -T\frac{\alpha}{2} p(i,j) + \frac{2}{q}\frac{\alpha}{2} p(i,j)(qT) - \frac{2}{q}\gamma_w(i,j)\left(\frac{qT}{2}\right) - \frac{2}{q}\gamma_b(i,j)\left(\frac{qT}{2}\right) \\
&= 0
\end{aligned}$$

by (B.5) and (B.6). Either way,

$$f_t(i,j) - \frac{\alpha}{2} p(i,j) + \frac{2}{q}\frac{\alpha}{2} p(i,j)\mathbb{1}_{q,t} - \frac{2}{q}\gamma_w(i,j)\mathbb{1}_{q,t}(1 - \mathbb{1}_{\text{odd},t}) - \frac{2}{q}\gamma_b(i,j)\mathbb{1}_{q,t}\mathbb{1}_{\text{odd},t} = f_{t+1}(i,j),$$

and (B.19) shows that we have satisfied

$$\epsilon_{t+1} \le C_\epsilon (t+1)^{3/5} + f_{t+1}(i,j).$$

For the reverse inequality, analogous calculations will show that, assuming (B.10) for some $t \ge KT$, all $i, j$, then

$$\epsilon_{t+1} \ge C_\epsilon (t+1)^{3/5} + f_{t+1}(i,j)$$

for all $i + j > 0$.

Finally, we return to the case $i = j = 0$. Here, we use the same type of argument as before, but this time we have additional terms to include. Recall that $\hat{N}_t(-1,0) = \hat{N}_t(0,-1) = p(-1,0) = p(0,-1) = 0$ by definition for all $t$, and so

from (B.1) and (3.4),

$$\begin{aligned}
\epsilon_{t+1}(0,0) &= E[\hat{N}_{t+1}(0,0)] - \frac{\alpha}{2}(t+1)p(0,0) \\
&= \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t} + E[\Delta_t(0,0)] \\
&+ \left(1 - \mathbb{1}_{q,t}\frac{\delta}{\frac{1}{2}(q+\alpha\delta)t}\left(\frac{q_{with}}{2q}\mathbb{1}_{\text{odd},t} + \frac{q_{bet}}{2q}(1-\mathbb{1}_{\text{odd},t})\right)\right)E[\hat{N}_t(0,0)] \\
&- \frac{\alpha}{2}tp(0,0) + \left(\frac{2}{q}-1\right)\frac{\alpha}{2}p(0,0) - \frac{2}{q}\left(\frac{\alpha}{2} + \gamma_w(0,0) + \gamma_b(0,0)\right).
\end{aligned}$$

If $\mathbb{1}_{q,t} = 0$, then

$$\epsilon_{t+1}(0,0) = \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t} + E[\Delta_t(0,0)] + \epsilon_t(0,0)\frac{\alpha}{2}p(0,0).$$

If $\mathbb{1}_{q,t} = \mathbb{1}_{\text{odd},t} = 1$, then

$$\begin{aligned}
\epsilon_{t+1}(0,0) &= \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t} + E[\Delta_t(0,0)] + \left(1 - \frac{q_{with}}{2q}\frac{\delta}{\frac{1}{2}(q+\alpha\delta)t}\right)E[\hat{N}_t(0,0)] \\
&- \frac{\alpha}{2}tp(0,0) + \left(\frac{2}{q}-1\right)\frac{\alpha}{2}p(0,0) - \frac{2}{q}\left(\frac{\alpha}{2} + \gamma_w(0,0) + \gamma_b(0,0)\right) \\
&= \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t} + E[\Delta_t(0,0)] + \left(1 - \frac{q_{with}}{2q}\frac{\delta}{\frac{1}{2}(q+\alpha\delta)t}\right)\epsilon_t(0,0) \\
&+ \left(\frac{2}{q}-1\right)\frac{\alpha}{2}p(0,0) - \frac{2}{q}\left(\frac{\alpha}{2} + \gamma_b(0,0)\right).
\end{aligned}$$

If $\mathbb{1}_{q,t} = 1 - \mathbb{1}_{\text{odd},t} = 1$, then

$$\begin{aligned}
\epsilon_{t+1}(0,0) &= \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t} + E[\Delta_t(0,0)] + \left(1 - \frac{q_{bet}}{2q}\frac{\delta}{\frac{1}{2}(q+\alpha\delta)t}\right)\epsilon_t(0,0) \\
&+ \left(\frac{2}{q}-1\right)\frac{\alpha}{2}p(0,0) - \frac{2}{q}\left(\frac{\alpha}{2} + \gamma_w(0,0)\right).
\end{aligned}$$

Combining the induction hypothesis (B.9) with the above three cases, we have shown that

$$\begin{aligned}
\epsilon_{t+1}(0,0) \le C_\epsilon(t+1)^{3/5} + f_t(0,0) + \mathbb{1}_{\alpha,t}\mathbb{1}_{\text{odd},t} - \frac{2}{q}\frac{\alpha}{2}\mathbb{1}_{q,t} - \frac{\alpha}{2}p(0,0) \\
+ \frac{2}{q}\frac{\alpha}{2}p(0,0)\mathbb{1}_{q,t} - \frac{2}{q}\gamma_w(0,0)\mathbb{1}_{q,t}(1-\mathbb{1}_{\text{odd},t}) - \frac{2}{q}\gamma_b(0,0)\mathbb{1}_{q,t}\mathbb{1}_{\text{odd},t}.
\end{aligned} \tag{B.21}$$

Recall that for $i = j = 0$, (B.11) becomes

$$f_t(0,0) = \sum_{s=0}^{(t \bmod T)-1} \Bigg( \mathbb{1}_{\alpha,s}\mathbb{1}_{\text{odd},s} - \frac{2}{q}\frac{\alpha}{2}\mathbb{1}_{q,s} - \frac{\alpha}{2}p(0,0) \\ + \frac{2}{q}\frac{\alpha}{2}p(0,0)\mathbb{1}_{q,s} - \frac{2}{q}\gamma_w(0,0)\mathbb{1}_{q,s}(1-\mathbb{1}_{\text{odd},s}) - \frac{2}{q}\gamma_b(0,0)\mathbb{1}_{q,s}\mathbb{1}_{\text{odd},s} \Bigg). \tag{B.22}$$

Thus we can write (B.21) as

$$\epsilon_{t+1}(0,0) \le C_\epsilon(t+1)^{3/5} + \sum_{s=0}^{t \bmod T} \Bigg( \mathbb{1}_{\alpha,s}\mathbb{1}_{\text{odd},s} - \frac{2}{q}\frac{\alpha}{2}\mathbb{1}_{q,s} - \frac{\alpha}{2}p(0,0) \\ + \frac{2}{q}\frac{\alpha}{2}p(0,0)\mathbb{1}_{q,s} - \frac{2}{q}\gamma_w(0,0)\mathbb{1}_{q,s}(1-\mathbb{1}_{\text{odd},s}) - \frac{2}{q}\gamma_b(0,0)\mathbb{1}_{q,s}\mathbb{1}_{\text{odd},s} \Bigg). \tag{B.23}$$

If $(t+1) \bmod T > 0$, then $t \bmod T = ((t+1) \bmod T) - 1$, and therefore (B.22) implies

$$\epsilon_{t+1}(0,0) \le C_\epsilon(t+1)^{3/5} + f_{t+1}(0,0). \tag{B.24}$$

If $(t+1) \bmod T = 0$, then $t \bmod T = T - 1$, and (B.5), (B.6), (B.7) yield

$$\begin{aligned}
&\sum_{s=0}^{T-1} \Bigg( \mathbb{1}_{\alpha,s}\mathbb{1}_{\text{odd},s} - \frac{2}{q}\frac{\alpha}{2}\mathbb{1}_{q,s} - \frac{\alpha}{2}p(0,0) \\
&+ \frac{2}{q}\frac{\alpha}{2}p(0,0)\mathbb{1}_{q,s} - \frac{2}{q}\gamma_w(0,0)\mathbb{1}_{q,s}(1-\mathbb{1}_{\text{odd},s}) - \frac{2}{q}\gamma_b(0,0)\mathbb{1}_{q,s}\mathbb{1}_{\text{odd},s} \Bigg) \\
&\quad = \frac{\alpha T}{2} - \frac{\alpha}{2}Tp(0,0) - \frac{2}{q}\frac{\alpha}{2}(qT) + \frac{2}{q}\frac{\alpha}{2}p(0,0)(qT) - \frac{2}{q}(\gamma_w(0,0)+\gamma_b(0,0))\left(\frac{qT}{2}\right) \\
&\quad = \frac{\alpha}{2}Tp(0,0) - \frac{\alpha}{2}T - T\gamma_w(0,0) - T\gamma_b(0,0) \\
&\quad = T\left(\frac{\alpha}{2}p(0,0) - \frac{\alpha}{2}p(0,0)\right) = 0,
\end{aligned}$$

which proves (B.24) for the final case.

As before, we can make the same argument to prove the reverse inequality.

□

## BIBLIOGRAPHY


[1] E. Abbe, A. S. Bandeira, and G. Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, Jan 2016.

[2] Emmanuel Abbe. Community Detection and Stochastic Block Models: Recent Developments. Technical report, 2018.

[3] William Aiello, Fan Chung, and Linyuan Lu. *Random Evolution in Massive Graphs*, pages 97–122. Springer US, Boston, MA, 2002.

[4] Roy M. Anderson and Robert M. (Robert McCredie) May. *Infectious diseases of humans : dynamics and control*. Oxford University Press, 1991.

[5] Benjamin Armbruster and Ekkehard Beck. An elementary proof of convergence to the mean-field equations for an epidemic model. *IMA Journal of Applied Mathematics*, 82(1):152–157, feb 2017.

[6] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group formation in large social networks: Membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '06, pages 44–54, New York, NY, USA, 2006. ACM.

[7] S. Banerjee, A. Chatterjee, and S. Shakkottai. Epidemic thresholds with external agents. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pages 2202–2210, April 2014.

[8] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[9] Marc Barthlemy, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *Journal of Theoretical Biology*, 235(2):275 – 288, 2005.

[10] Laura M. Bogart, Mosepele Mosepele, Nthabiseng Phaladze, Bright Lekoko, David J. Klein, Sarah MacCarthy, and Harold D. Green. A Social Network Analysis of HIV Treatment Partners and Patient Viral Suppression in Botswana. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 78(2):183–192, jun 2018.

[11] Marián Boguñá and Romualdo Pastor-Satorras. Epidemic spreading in correlated complex networks. *Phys. Rev. E*, 66:047104, Oct 2002.

[12] Béla Bollobás, Christian Borgs, Jennifer Chayes, and Oliver Riordan. Directed scale-free graphs. In *Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '03, pages 132–139, Philadelphia, PA, USA, 2003. Society for Industrial and Applied Mathematics.

[13] Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures and Algorithms*, 31(1):3–122, aug 2007.

[14] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Structures & Algorithms*, 18(3):279–290, 2001.

[15] T. N. Bui, S. Chaudhuri, F. T. Leighton, and M. Sipser. Graph bisection algorithms with good average case behavior. *Combinatorica*, 7(2):171–191, jun 1987.

[16] D. Buraczewski, E. Damek, T. Mikosch, and J. Zienkiewicz. Large deviations for solutions to stochastic recurrence equations under kesten's condition. *The Annals of Probability*, 41(4):2755–2790, 2013.

[17] Neal Caren. A sociology citation network, 2012. http://nealcaren.web.unc.edu/a-sociology-citation-network/.

[18] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 675–684, New York, NY, USA, 2011. ACM.

[19] Xiangwei Chu, Jihong Guan, Zhongzhi Zhang, and Shuigeng Zhou. Epidemic spreading in weighted scale-free networks with community structure. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(07):P07043, jul 2009.

[20] A. Clauset, C. Shalizi, and M. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.

[21] Aaron Clauset, M. E. J. Newman, and Cristopher Moore. Finding community structure in very large networks. *Phys. Rev. E*, 70:066111, Dec 2004.

[22] S. Coles, J. Heffernan, and J. Tawn. Dependence measures for extreme value analyses. *Extremes*, 2:339–365, 1999.

[23] Colin Cooper and Alan Frieze. A general model of web graphs. *Random Structures & Algorithms*, 22(3):311–335, 2003.

[24] J.-J. Daudin, F. Picard, and S. Robin. A mixture model for random graphs. *Statistics and Computing*, 18(2):173–183, Jun 2008.

[25] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E*, 84:066106, Dec 2011.

[26] Güven Demirel, Edmund Barter, and Thilo Gross. Dynamics of epidemic diseases on a growing adaptive network. *Scientific Reports*, 7:42352, Feb 2017.

[27] Sergey Dorogovtsev and Jos Fernando Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW*, volume 57. 03 2003.

[28] M.E Dyer and A.M Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451–489, dec 1989.

[29] David Ediger, Karl Jiang, Jason Riedy, David A. Bader, and Courtney Corley. Massive Social Network Analysis: Mining Twitter for Social Good. In *2010 39th International Conference on Parallel Processing*, pages 583–593. IEEE, sep 2010.

[30] Santo Fortunato and Marc Barthélemy. Resolution limit in community detection. Technical Report 1, 2007.

[31] Santo Fortunato and Darko Hric. Community detection in networks: A user guide. *Physics Reports*, 659:1 – 44, 2016. Community detection in networks: A user guide.

[32] Madhavi K Ganapathiraju, Mohamed Thahir, Adam Handen, Saumendra N Sarkar, Robert A Sweet, Vishwajit L Nimgaonkar, Christine E Loscher, Eileen M Bauer, and Srilakshmi Chaparala. Schizophrenia interactome with 504 novel protein-protein interactions. *NPJ schizophrenia*, 2:16012, 2016.

[33] A. Ganesh, L. Massoulie, and D. Towsley. The effect of network topology on the spread of epidemics. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 2, pages 1455–1466 vol. 2, March 2005.

[34] Johannes Gehrke, Paul Ginsparg, and Jon Kleinberg. Overview of the 2003 KDD Cup. Technical report.

[35] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.

[36] Beniamino Guerra and Jesús Gómez-Gardeñes. Annealed and mean-field formulations of disease dynamics on static and adaptive networks. *Phys. Rev. E*, 82:035101, Sep 2010.

[37] S Gupta, R M Anderson, and R M May. Networks of sexual contacts: implications for the pattern of spread of HIV. *AIDS (London, England)*, 3(12):807–17, dec 1989.

[38] Bruce Hajek and Suryanarayana Sankagiri. Recovering a Hidden Community in a Preferential Attachment Graph. In *IEEE International Symposium on Information Theory - Proceedings*, volume 2018-June, pages 2097–2101, 2018.

[39] R. van der Hofstad. *Random Graphs and Complex Networks. Volume 1.* Cambridge Series in Statistical and Probabilistic Mathematics, 2017.

[40] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social Networks*, 5(2):109 – 137, 1983.

[41] Petter Holme and Beom Jun Kim. Growing scale-free networks with tunable clustering. *Phys. Rev. E*, 65:026107, Jan 2002.

[42] Petter Holme and Jari Saramäki. Temporal networks. *Physics Reports*, 519(3):97 – 125, 2012. Temporal Networks.

[43] D. Hric, K. Kaski, and M. Kivelä. Stochastic Block Model Reveals the Map of Citation Patterns and Their Evolution in Time. *ArXiv e-prints*, April 2017.

[44] Wei Huang and Chunguang Li. Epidemic spreading in scale-free networks with community structure. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(01):P01014, 2007.

[45] Bernardo A. Huberman, Daniel M. Romero, and Fang Wu. Social networks that matter: Twitter under the microscope. *CoRR*, abs/0812.1045, 2008.

[46] Jonathan Jordan. Geometric preferential attachment in non-uniform metric spaces. *Electronic Journal of Probability*, 18(0), 2013.

[47] Brian Karrer and M. E. J. Newman. Stochastic blockmodels and community structure in networks. *Phys. Rev. E*, 83:016107, Jan 2011.

[48] István Z. Kiss, Joel C. Miller, and Péter L. Simon. *Mathematics of Epidemics on Networks*, volume 46 of *Interdisciplinary Applied Mathematics*. Springer International Publishing, Cham, 2017.

[49] A S Klovdahl, J J Potterat, D E Woodhouse, J B Muth, S Q Muth, and W W Darrow. Social networks and infectious disease: the Colorado Springs Study. *Social science & medicine (1982)*, 38(1):79–88, jan 1994.

[50] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of growing random networks. *Phys. Rev. Lett.*, 85:4629–4632, Nov 2000.

[51] Ravi Kumar, Jasmine Novak, and Andrew Tomkins. *Structure and Evolution of Online Social Networks*, pages 337–357. Springer New York, New York, NY, 2010.

[52] Thomas G. Kurtz. Solutions of ordinary differential equations as limits of pure jump markov processes. *Journal of Applied Probability*, 7(01):49–58, apr 1970.

[53] Thomas G. Kurtz. Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8(02):344–356, jun 1971.

[54] Carl A Latkin, Melissa A Davey-Rothwell, Amy R Knowlton, Kamila A Alexander, Chyvette T Williams, and Basmattee Boodram. Social network approaches to recruitment, HIV prevention, medical care, and medication adherence. *Journal of acquired immune deficiency syndromes (1999)*, 63 Suppl 1(0 1):S54–8, jun 2013.

[55] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time. In *Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining - KDD '05*, page 177, New York, New York, USA, 2005. ACM Press.

[56] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graph evolution. *ACM Transactions on Knowledge Discovery from Data*, 1(1):2–es, mar 2007.

[57] Chanchan Li, Guo ping Jiang, Yurong Song, Lingling Xia, Yinwei Li, and Bo Song. Modeling and analysis of epidemic spreading on community networks with heterogeneity. *Journal of Parallel and Distributed Computing*, 119:136 – 145, 2018.

[58] Zonghua Liu and Bambi Hu. Epidemic spreading in community networks. *Europhys. Lett*, 72(2):315–321, 2005.

[59] Shan Lu, G. Samorodnitsky, W. Gong, B. Jiang, J. Kang, and D. Towsley. Poisson process driven stochastic differential equations for bivariate heavy tailed distributions. In *2016 American Control Conference (ACC)*, pages 2977–2982, July 2016.

[60] C R Maccluer. The Many Proofs and Applications of Perron's Theorem *. Technical Report 3, 2000.

[61] S. Meyn and R. Tweedie. Stability of markovian processes ii: Continuous-time processes and sampled chains. *Applied Probability Trust*, 25:487–517, 1993.

[62] Matthieu Nadini, Kaiyuan Sun, Enrico Ubaldi, Michele Starnini, Alessandro Rizzo, and Nicola Perra. Epidemic spreading in modular time-varying networks. *Scientific Reports*, 8(1):2352, dec 2018.

[63] M. E. J. Newman. Spread of epidemic disease on networks. *Phys. Rev. E*, 66:016128, Jul 2002.

[64] M. E. J. Newman. Communities, modules and large-scale structure in networks. *Nature Physics*, 8:25–31, 2012.

[65] Roberto Oliveira and Joel Spencer. Connectivity transitions in networks with super-linear preferential attachment. *Internet Math.*, 2(2):121–163, 2005.

[66] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435:814–818, Jun 2005.

[67] Zizi Papacharissi. The virtual geographies of social networks: a comparative analysis of facebook, linkedin and asmallworld. *New Media & Society*, 11(1-2):199–220, 2009.

[68] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87:925–979, Aug 2015.

[69] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3203, Apr 2001.

[70] Xiao-Long Peng, Michael Small, Xin-Jian Xu, and Xinchu Fu. Temporal prediction of epidemic patterns in community networks. *New Journal of Physics*, 15(11):113033, nov 2013.

[71] Edwin Perkins, Stewart N. Ethier, and Thomas G. Kurtz. Markov Processes, Characterization and Convergence. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 151(2):367, 2006.

[72] Clara Pizzuti. Evolutionary Computation for Community Detection in Networks: A Review. *IEEE Transactions on Evolutionary Computation*, 22(3):464–483, jun 2018.

[73] Pascal Pons and Matthieu Latapy. Computing Communities in Large Networks Using Random Walks. pages 284–293. Springer, Berlin, Heidelberg, oct 2005.

[74] Maoying Qiao, Jun Yu, Wei Bian, Qiang Li, and Dacheng Tao. Improving stochastic block models by incorporating power-law degree characteristic. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 2620–2626, 2017.

[75] William J. Reed and Barry D. Hughes. From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. *Phys. Rev. E*, 66:067103, Dec 2002.

[76] R B Rothenberg, J J Potterat, D E Woodhouse, S Q Muth, W W Darrow, and A S Klovdahl. Social network dynamics and HIV transmission. *AIDS (London, England)*, 12(12):1529–36, aug 1998.

[77] Marcel Salathé and James H. Jones. Dynamics and Control of Diseases

in Networks with Community Structure. *PLoS Computational Biology*, 6(4):e1000736, apr 2010.

[78] G. Samorodnitsky. *Stochastic Processes and Long Range Dependence*. Operations Research and Financial Engineering. Springer, Switzerland, 2016.

[79] Gennady Samorodnitsky, Sidney Resnick, Don Towsley, Richard Davis, Amy Willis, and Phyllis Wan. Nonstandard regular variation of in-degree and out-degree in the preferential attachment model. *J. Appl. Probab.*, 53(1):146–161, 03 2016.

[80] Péter L. Simon, Michael Taylor, and Istvan Z. Kiss. Exact epidemic models on graphs using graph-automorphism driven lumping. *Journal of Mathematical Biology*, 62(4):479–508, Apr 2011.

[81] R. D. Smith. Instant Messaging as a Scale-Free Network. *eprint arXiv:cond-mat/0206378*, June 2002.

[82] Bo Song, Yu-Rong Song, and Guo-Ping Jiang. How clustering affects epidemics in complex networks. In *2017 International Conference on Computing, Networking and Communications (ICNC)*, pages 178–183, Jan 2017.

[83] Clara Stegehuis, Remco van der Hofstad, and Johan S. H. van Leeuwaarden. Epidemic spreading on complex networks with community structures. *Scientific Reports*, 6(29748), 2016.

[84] E. C. (Edward Charles) Titchmarsh. *Introduction to the theory of Fourier integrals*. Chelsea Pub. Co, New York N.Y., 3rd ed. edition, 1986.

[85] Eugenio Valdano, Michele Re Fiorentin, Chiara Poletto, and Vittoria Colizza. Epidemic threshold in continuous-time evolving networks. *Phys. Rev. Lett.*, 120:068302, Feb 2018.

[86] Piet Van Mieghem. Decay towards the overall-healthy state in SIS epidemics on networks. oct 2013.

[87] Piet Van Mieghem, Dragan Stevanović, Fernando Kuipers, Cong Li, Ruud van de Bovenkamp, Daijie Liu, and Huijuan Wang. Decreasing the spectral radius of a graph by link removals. *Phys. Rev. E*, 84:016101, Jul 2011.

[88] Wei Wang, Ming Tang, H Eugene Stanley, and Lidia A Braunstein. Uni-

fication of theoretical approaches for epidemic spreading on complex networks. *Reports on Progress in Physics*, 80(3):036603, 2017.

[89] Xiaoyan Wu and Zonghua Liu. How community structure influences epidemic spread in social networks. *Physica A: Statistical Mechanics and its Applications*, 387(2):623 – 630, 2008.

[90] Xu Xin-Ping, Liu Feng, and Li Wei. Growing small-world networks based on a modified ba model. *Chinese Physics Letters*, 23(3):750, 2006.

[91] Jaewon Yang and Jure Leskovec. Defining and evaluating network communities based on ground-truth. *Knowledge and Information Systems*, 42(1):181–213, jan 2015.

[92] Hao Yin, Austin R. Benson, Jure Leskovec, and David F. Gleich. Local Higher-Order Graph Clustering. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17*, pages 555–564, New York, New York, USA, 2017. ACM Press.

[93] Yunpeng Zhao, Elizaveta Levina, and Ji Zhu. Consistency of community detection in networks under degree-corrected stochastic block models. *Ann. Statist.*, 40(4):2266–2292, 08 2012.